Evaluation of Liquid Biopsy Workflows and *ESR1* Alterations in Breast Cancer

Silvia Rita Vitale

The studies described in this thesis were performed within the framework of the Erasmus Postgraduate School of Molecular Medicine at the department of Medical Oncology and Cancer Genomics Netherlands, Erasmus MC – Cancer Institute, Erasmus University Medical Center, Rotterdam, The Netherlands.

The research project were supported in part by: *i*. KWF-Alpe d'HuZes projects [EMCR 2014-6340 and NKI 2014- 7080], and a grant from Cancer Genomics Netherlands (CGC.nl)/Netherlands Organization for Scientific Research (NWO) (*Chapter 3, 5, 6 and 7*), *ii*. MRace pilot grant 2012 (*Chapter 4*), *iii*. European Research Council advanced grant 322737 (*Chapter 6 and 7*), and *iv*.a Pink Ribbon Netherlands grant (project WO 61) (*Chapter 6*).

Support for the printing of this thesis was obtained by kind contribution from:

Department of Medical Oncology, Erasmus Medical Center

Erasmus University Rotterdam

Cover design by Sabrina Arena

Lay-out by Silvia Rita Vitale

Printed by GILDEPrint, Javastraat 123, Enschede, OV 7512 ZE, Netherlands

ISBN/EAN: 978-94-6419-505-7

Evaluation of Liquid Biopsy Workflows and *ESR1* Alterations in Breast Cancer

Evaluatie van Vloeistofbiopt Werkschema's en *ESR1* Veranderingen in Borstkanker

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Wednesday 29 June 2022at 10:30hrs
By
Silvia Rita Vitale
born in Catania, Italy.

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	prof. dr. J.W.M. Martens
Other members:	dr. M. J. Hooning prof. dr. E.M.D. Schuuring prof. dr. S. Van Laere
Copromotor:	dr. ir. M.P.H.M. Jansen

Paranimfen:	J.C.A. Helmijr dr. Stefania Stella

So far so good…

To my best friend

Anieta

Contents

Chapter 1 General Introduction **9**

Liquid biopsies in solid cancers

Cell-free DNA and circulating tumor DNA

Circulating extracellular vesicles

Digital PCR and next generation sequencing

Current challenges in liquid biopsy research

The role of Estrogen Receptor in breast cancer

Monitoring variation in *ESR1* in endocrine resistance

- *ESR1* Mutations
- *ESR1* Splice variants
- *ESR1* fusions
- *ESR1* methylation

Chapter 2 Aims and outline of this thesis **31**

Chapter 3 High-throughput isolation of circulating tumor DNA: a comparison of automated platforms

Vitale SR*, van Dessel LF*, Helmijr JCA, Wilting SM, van der Vlugt-Daane M, Oomen-de Hoop E, Sleijfer S, Martens JWM, Jansen MPHM, Lolkema MP

Mol Oncol. 2019 Feb; 13(2):392-402. doi: 10.1002/1878-0261.12415 **33**

Chapter 4 *TP53* mutations in serum circulating cell-free tumor DNA as longitudinal biomarker for high-grade serous ovarian cancer

Vitale SR, Groenendijk FH, van Marion R, Beaufort C, Helmijr J, Jan Dubbink H, Dinjens WNM, Ewing-Graham PC, Smolders R, van Doorn LHC, Boere IA, Berns EMJJ, Helleman J, Jansen MPHM

Biomolecules 2020 Mar 7; 10(3):415. doi: 10.3390/biom10030415 **59**

Chapter 5 **Detection of tumor-derived extracellular vesicles in plasma from patients with solid cancer**

Vitale SR*, Jean A. Helmijr JA*, Gerritsen M, Coban H, van Dessel LF, Beije N, van der Vlugt-Daane M, Vigneri P, Sieuwerts AM, Dits N, van Royen ME, Jenster G, Sleijfer S, Lolkema M, Martens JWM, P.H.M. Jansen MPHM *

BMC Cancer 21, 315 (2021). https://doi.org/10.1186/s12885-021-08007-z **85**

Chapter 6 **An Optimized Workflow to Evaluate Estrogen Receptor Gene Mutations in Small Amounts of Cell-Free DNA**

Vitale SR*, Sieuwerts AM*, Beije N, Kraan J, Angus L, Mostert B, Reijm EA, Van NM, van Marion R, Dirix LY, Hamberg P, de Jongh FE, Jager A, Foekens JA, Vigneri P, Sleijfer S, Jansen MPHM, Martens JWM

J Mol Diagn. 2019 Jan;21(1):123-137. doi: 10.1016/j.jmoldx.2018.08.010. Epub 2018 Oct 6 **129**

Chapter 7 **The prognostic and predictive value of *ESR1* fusion gene transcripts in primary breast cancer**

Vitale SR, Ruigrok-Ritstier K, Timmermans AM, Foekens R, Trapman-Jansen AMA., Beaufort CM, Vigneri P, Sleijfer S,. Martens JWM, Sieuwerts AM, Jansen MPHM,

BMC Cancer 22, 165 (2022). doi: 10.1186/s12885-022-09265-1 **177**

Chapter 8 **General Discussion** **217**

Chapter 9 **English Summary/Samenvatting/Sommario** **237**

Appendices **251**

PhD Portfolio **252**

List of publications **255**

Curriculum Vitae **261**

Acknowledgments **263**

**These authors contributed equally to the manuscript*

Chapter 1

General Introduction

General Introduction

Over the past decades, cancer has become the second leading cause of death in our society. To address this, personalized medicine has become one of the most studied areas to improve cancer therapy efficacy [1]. As a result, the patient's outcome has improved due to a more accurate diagnosis allowing more effective therapeutic interventions. Additionally, and importantly, side effects and unnecessary treatment costs are avoided [2].

In clinical practice, evaluating a primary tumor tissue biopsy is the routine approach to diagnose cancer in patients. Furthermore, molecular characteristics assessed in this specimen are most commonly used to stratify individual patients guiding the oncologists on therapy choice [2]. Since tissue biopsies are usually taken from the primary tumor, they reflect its molecular composition at time of diagnosis of the disease. However, the disease' characteristics may change over time during progression for the selection pressure on the heterogeneous tumor [3]. Moreover, taking biopsies is not always feasible, especially if multiple biopsies are needed or if lesions are inaccessible. Due to their invasive nature such procedures are generally considered patient-unfriendly, risky and costly. One other limitation is that a single biopsy may not be representative due to the fact that the intra-patient heterogeneity between individual lesions is not properly sampled. As a result, techniques currently used to sample a lesion might overlook small aggressive subclones, making this approach less optimal for treatment decisions. Moreover, most biopsies are preserved by formaldehyde fixation and paraffin embedding (FFPE), a procedure not ideal for downstream molecular analyses. Liquid biopsies, which allow studying tumor cells or tumor cell content circulating in body fluids (blood, urine, saliva, ascites, liquor) might alleviate the issues raised above. Currently, blood biopsies are particularly explored for solid cancer diagnostics as those can be less invasively and thus more easily obtained, also allowing sequential monitoring of disease evolution [2, 4-6].

Liquid Biopsies in solid cancers

Liquid biopsies provide an alternative when traditional tumor biopsies are impossible or difficult to obtain and may even provide additional clinically useful information not present in a single solid tumor biopsy [7] (**Figure 1**). As the procedure only requires a simple blood draw, additional advantages are the rapid procedure and cost effectiveness for taking a sample. Moreover, blood samples can repeatedly be taken during the course of a patient's treatment. Consequentially,

changes, for example in the quantity and/or characteristics of tumor derived DNA within the blood, can indicate whether a particular therapy is effective, or whether a tumor is evolving towards drug-resistance, indicating the need for a treatment switch [5].

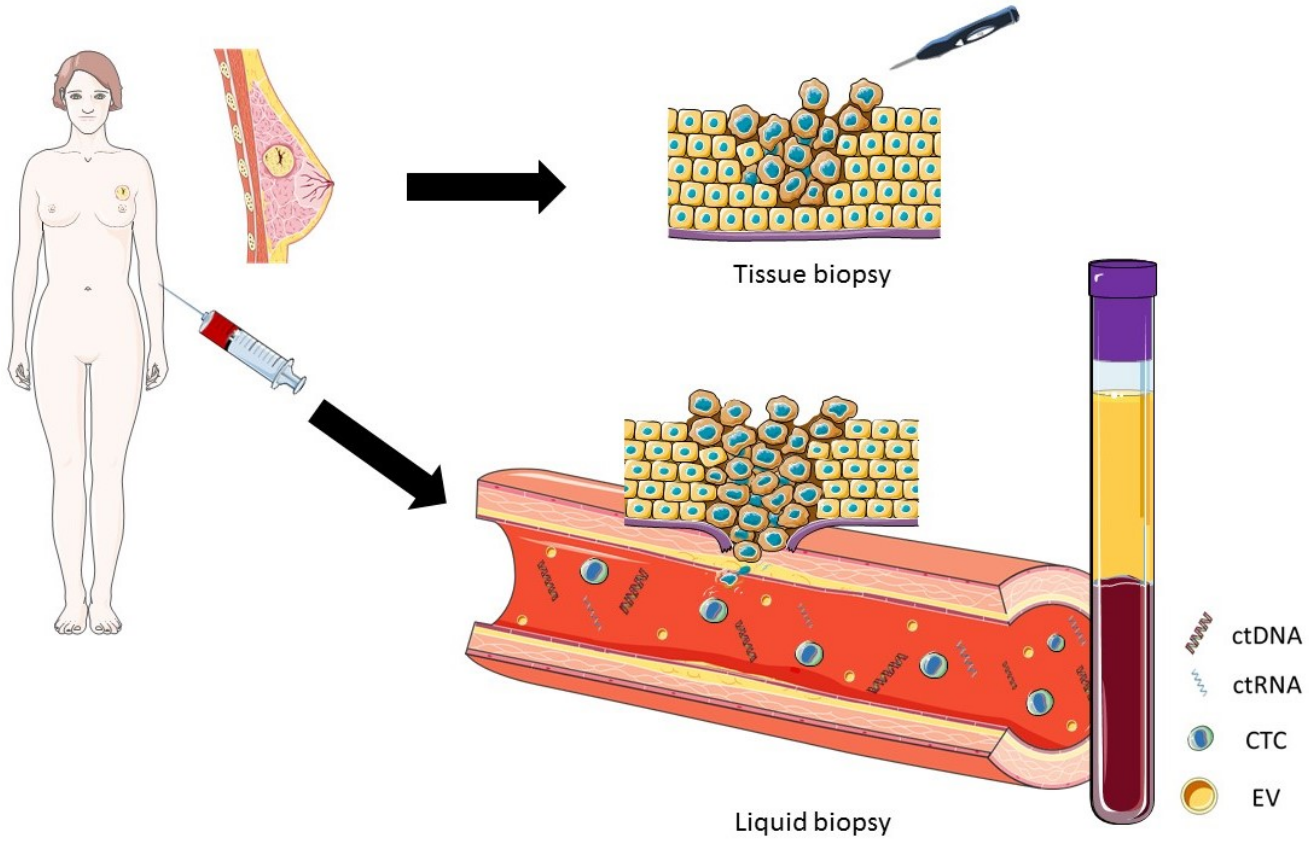


Figure 1. Tissue biopsy *versus* Liquid biopsy

Cell-free DNA and circulating tumor DNA

Cell-free DNA (cfDNA) is composed of deoxynucleic acid fragments that circulate outside of cells in bodily fluids (e.g. plasma, serum, urine) and which is released by tumor cells but also by normal cells as part of ordinary cell physiological processes (e.g. apoptosis and necrosis). After its discovery in 1948 by *Mandel and Metais* [8], cfDNA has become recognized as a very promising marker for the diagnosis and prognosis of cancer patients [9]. Importantly, studies reported that higher amounts of cfDNA could be isolated from cancer patients compared to healthy individuals [10]. Additionally, cfDNA concentration was shown to be related with tumor burden, but is also influenced by various processes such as physical exercise, inflammation and cellular turnover [11]. The fact that both normal and tumor related processes generate cfDNA makes it difficult to distinguish within cfDNA the DNA that is derived from tumor cells, i.e. cell free tumor DNA (ctDNA), from those originating from normal cells [9]. Only tumor-specific (somatic) genomic aberrations will discriminate between cfDNA derived from tumor cells and those from normal cells [9, 12]. Knowledge about somatically acquired genomic aberrations in cancer have been gained in the last decade by international consortia that performed whole genome and exome sequencing (WGS, WES) on large cohorts of tumor tissues from multiple cancer types. This results in repositories like Catalogue of Somatic Mutation In Cancer (COSMIC) that report for each cancer-type the tumor-associated genetic alterations, such as single nucleotide variants (SNVs), copy number alterations (CNA), and structural variants (SVs) [13] 14]. These tumor-specific DNA alterations can be used to distinguish ctDNA from normal cfDNA. The ctDNA often represents only 0.1-10% of the total cfDNA [11, 15-17]. However, the ctDNA load depends on the stage of disease, cancer type, timing of blood draw and type and particularly effectiveness of a particular therapeutic intervention (surgery, radiotherapy and/or systemic therapy) but also on pre-and post-analytical conditions of the blood biopsy taken [18].

Many previous reports have shown the feasibility of molecular characterization of cfDNA [12, 19], but still many challenges remain due to its extremely low concentration (frequently below 10 ng/mL) and fragmented nature (sizes comprise between 140-170 bp) as well as to the fact that a minority of the material being actually tumor derived [4]. For instance, ctDNA levels can be diluted by wild type DNA released from lysed hematological white blood cells that may result in an underestimation of the ctDNA fraction which can drop even below the detection limit of the analysis method [9]. As described in this thesis and elsewhere [20, 21], this phenomenon can be minimized by reducing time between blood draw and plasma collection and by choosing

the correct type of preservative in blood collection tubes.

Circulating extracellular vesicles

In the past decade, extracellular vesicles (EVs) have emerged as important mediators of intercellular communications. EVs consist of small membrane containing vesicles shedded into the extracellular space by both normal and tumor cells [22, 23]. EVs are produced by invagination of the lipid bilayer, as for example stimulated by plasma membrane activation through intracellular calcium influx [24]. EVs include different subtypes of vesicles such as micro-vesicles (MVs), exosomes and apoptotic bodies that have been characterized based on their biogenesis or release pathways.

Micro-vesicles (MVs) (size 100 nm) are buddings from the plasma membrane and contain cytoplasmic cargo [25]. Exosomes (size ranging from 40 to 120 nm) are derived from the fusion between multi-vesicular bodies and the plasma membrane [26,27]. Apoptotic bodies (size 50 nm–2 μm) originating from dying cells, can be more abundant than exosomes or MVs under specific conditions and can vary in content between biofluids [28]. Currently, many (commercially) isolation methods for EVs are not able to distinguish between these subtypes, therefore I used the term EVs in this thesis to refer to all these different types. EVs circulate in different body fluids, such as blood and urine. They are composed of a lipid bilayer and their cargo contains proteins including transcription factors, messenger RNA (mRNA), microRNA (miRNA), DNA and lipids, suggesting they can mediate signaling to near but also distant cells [23]. Actually, cancer EVs might transfer pathogenic components not only to their surrounding cells but also to distant microenvironments, stimulating niche formation supporting of metastatic tumor cells. Circulating tumor EVs (ctEVs) can promote cell proliferation and migration, enhance angiogenesis and promote the development of metastases [29-31]. EVs derived from cancer cells can also stimulate the variation of host cell phenotypes to facilitate tumor growth, as already demonstrated *in vitro* in EVs derived from ovarian[32], breast[33] and renal cancer cells [31]. As a result, the isolation, quantification and characterization of circulating EVs might represent a promising and attractive clinical tool to characterize features of tumor cells. Actually, EVs, compared to other molecular biomarkers, show several advantages such as the expression of specific surface proteins in their membrane or the presence of tumor-specific RNA that might mirror their parental cells. The amount of RNA transported by EVs (EV-RNA) differs depending on the cell type of origin[23]. Studies reported that some cancer-derived EVs contain more total RNA

than those derived from normal cells [34]. Additionally, although they contain many transcripts in common with those of parental cells, some RNA might be systematically enriched in the released EVs [24, 29]. Therefore, the detection of target mutations in EV-RNA, as well as the expression of fusion genes and splice variants, emerge as an attractive field in clinical practice [35]. However, due to the low amount of RNA recovered from EVs, it requires the use of very sensitive methods for their isolation and characterization.

Digital PCR and next generation sequencing

Circulating biomarkers have been increasingly used in assessing tumor characteristics, which is particularly critical when a biopsy cannot be properly acquired or when repeated sampling is necessary [36]. This clinical scenario requires the use of sensitive and accurate techniques, such as digital PCR (dPCR) and Next Generation Sequencing (NGS) approaches [37, 38]. Several studies have investigated the potential of dPCR to identify ctDNA in plasma of patients with different tumor types and to monitor in time the response to therapies and/or to predict the progression of disease [39-42]. The dPCR method is based on the principle of diluting DNA into large numbers of individual nano-scale compartments, allowing single molecule analyses and, as a result, the detection of rare mutant molecules among common wild-type copies by amplification of individual DNA molecules [6, 43]. Different platforms have been developed for dPCR. The first systems, referred as first-generation digital PCR, was BEAMing and has been applied for molecular screening in different types of cancer, including breast, gastrointestinal carcinoma, lung and colorectal cancer (CRC) [6, 11]. The detection limit of this procedure is one mutant DNA molecule in a background of 10,000 wild-type molecules [38]. However, the first-generation technology required relatively cumbersome and quite complicated procedures, not suitable for routine clinical use. Second-generation dPCR alternatives, replacing BEAMing, included chip-based digital PCR Systems [Quantstudio 12k/3D (Thermo Fisher Scientific), Naica (Stilla Technologies)] and droplet-based digital PCR (ddPCR) Systems [QX-100/QX-200 (Biorad)] [38, 44]. In such systems, cfDNA is partitioned into many individual and parallel PCR reactions acting as independent micro-compartments. Each of these reactions ideally contain no more than one molecule and all reagents to amplify the target mutant sequence and the other one wild-type allele. Altogether, these methods allow the identification of mutant sequences within a wild-type background with a detection limit below 0.001% [38]. All systems allow the analysis of DNA extracted from clinical samples, such as FFPE tissues, or liquid biopsy samples. Molecular

biomarkers are measurable and detectable at both the DNA and RNA level by such dPCR methodologies also allowing tracing known mutations over time. For example, dPCR is commonly used in routine diagnostics for the detection of *EGFR* mutations in cfDNA derived from blood which is particularly relevant for clinical management of non-small lung cancer (NSCLC) patients treated with EGFR directed monoclonal antibodies [37, 38, 43]. An interesting study by *Misasale et al*. showed that *KRAS* mutations are associated to acquire resistance to anti-EGFR treatment in colorectal cancer patients. In this paper, the authors were able to measure the *KRAS* mutations by using patient blood samples and the BEAMing approach [45]. Similarly, dPCR has been used to assess the development of resistance mechanisms, as in the case of *ESR1* mutations in ER-positive metastatic breast cancer patients treated with endocrine therapies [44, 46-48]. Additionally, *PIK3CA* mutations determined by dPCR in cfDNA can be used to predict good prognosis in triple-negative breast cancer patients [49]. Other multiplex strategies have been developed to screen for a pool of mutations such as *RAS/RAF* mutations and *EGFR* exon 19 deletions [38]. Digital PCR has also been employed to evaluate circulating cell free RNA and miRNA levels derived from blood in different types of malignancies, including melanoma, lung, and colorectal carcinoma, and to evaluate ribonucleic acids from exosomes, such as the androgen receptor splice variant 7 (*AR-V7*) which is of clinical relevance in castration-resistant prostate cancer (CRPC) [41, 50, 51].

NGS is a high throughput process that permits the sequencing of millions of DNA molecules at the same time. Sequence reads of these molecules are then aligned to a reference genome sequence in order to identify genetic or epigenetic changes. Compared to traditional methods (e.g. Sanger sequencing), NGS offers advantages particularly in sensitivity and speed. In fact, standard NGS techniques have become the gold standard for the detection of mutations. Currently, different targeted NGS panels have been developed for detection of (hotspot) mutations in cancer-type relevant genes.

The NGS wide application is however still limited due to the relatively high costs and low sensitivity, the latter particularly important for the detection of rare mutations in liquid biopsies. Compared to standard NGS, dPCR is easy and fast, does not require complex informatics support for analysis and is superior in sensitivity and feasibility [38]. However, dPCR requires prior knowledge regarding the mutation to be monitored and the use of highly allele-specific probes to reduce cross-reactivity and false positives [37]. On the contrary, NGS can assess multiple genes in a single assay and has the capacity to identify novel alterations. Moreover, the introduction of

molecular barcodes to uniquely tag single DNA molecules reduce errors due to background noise and increase the capability to detect target sequences although at low levels [52]. However, there are still some technical limitations that have to be overcome to finalize the application of this approach into clinical practice for liquid biopsies. Nowadays, many strategies have combined the use of the two methodologies for liquid biopsy analysis, reaching a limit of detection of ctDNA as low as 0.001% depending on input amounts [38, 53, 54]. However, it has been noted that the minute amount of ctDNA released into circulation can limit the application of these type of analyses.

Current challenges in liquid biopsy research

The potential of liquid biopsies in personalized medicine has been highlighted by several studies, which have shown its feasibility to track tumor heterogeneity as well as the development of distant metastases or recurrence disease [13]. However, some serious challenges need to be resolved before liquid biopsies can be included in routine diagnostic settings. Firstly, implementation of liquid biopsy-based detection requires the standardization of workflows for both pre- and post-analytical phases. Specifically, the standardization of collection tube use for each biofluid, centrifugation steps, isolation methods and storage conditions as well as analytical steps, e.g. quantitation and data analysis methods, performed to guarantee reproducible data. Additionally, as previously reported, biofluid components might be influenced by physiological and lifestyle factors, such as age, stress and other behaviors, and by cellular turnover [4]. This means that the composition of plasma and/or serum components need to be investigated in healthy individuals before it is transposed to cancer's patients. Likewise, an in-depth analysis of the biology and mechanisms involved in the release of such components will help us to better understand if the molecular profile captured in the liquid biopsies really reflect the disease.

It has already been demonstrated that ctDNA is a promising biomarker for personalized medicine [55-57]. However, the principal hurdle to improve its clinical use, involves the limited number of available tumor-specific genomic aberrations to discriminate between normal and tumor cfDNA. Moreover, the amount of ctDNA within total cfDNA (ranging from <1% to >50%) is strongly correlated with both tumor burden and stage of disease. Hence, liquid biopsies compulsorily requires the use of sensitive, accurate and precise methods that correctly assess the origin of ctDNA traceable in body fluids. Moreover, as levels of mutant molecules vary greatly among

patients and can occurs as low as 0.01% of cfDNA in patients with early-stage disease [57], it is mandatory to increase the limit of detection by developing methods with high sensitivity and specificity for detection of such rare mutations. A possible improvement might be achieved by adding appropriate internal and external quality controls to each assay as well as the application of molecular barcodes, the latter which is currently already frequently applied [57].

Ideally, liquid biopsy would employ ctDNA, EVs and ctRNA all together to provide information about primary tumors and/or local or distant metastatic disease characteristics [55]. However, tumor heterogeneity is still a remaining issue in the field of liquid biopsies in clinical oncology, which demands future studies to gain better insight if such biomarkers actually offer a full or proper representation of the disease in situ. Moreover, it still has to be determined whether all metastases equally contribute to these biomarkers in the bloodstream. Another question that needs to be addressed relates to the different information that can be obtained from these liquid biopsy biomarkers as they have different origin. For example, the data gained from EVs and cfRNA that originates from living cells might be different from findings derived from ctDNA that is considered to be derived from dying/apoptotic cells.

It has also been demonstrated that technologies utilizing ctDNA are able to reveal the presence of tumor lesions or a clinical relapse earlier than (standard) imaging systems (e.g. CT and PET scanning) or before clinical signs arise [9, 57]. In this context, analyzing a liquid biopsy, which is cheaper compared to these standard approaches, would help to follow patients over time and allow identification of subjects who might benefit from a specific treatment or identify early-on individuals who has failed an existing treatment. A previous study has shown that persistent presence of ctDNA implies an inferior recurrence-free survival in patients with stage II colon cancer treated with adjuvant chemotherapy [58]. Similarly, postoperative measurement of ctDNA in plasma derived from NSCLC patients might reveal resistance to adjuvant chemotherapy [19]. Although this is a promising aspect of liquid biopsies, no clinical guidelines are available yet for its use in sample monitoring. Therefore, it is difficult to establish whether a mutation detected in ctDNA would actually affect the patient's management, particularly when measured at low level. This point might become an important issue when standard imaging does not show evidence for relapse. An additional issue might be the presence of somatic mutations in normal tissue [59, 60]. To solve this issue, we firstly need to establish a baseline value for each of these mutations in healthy tissue.

Finally, the study of biofluids-RNA might provide information about the temporal dynamics of the tumor allowing analyses of fusion genes as well as expression profiling of relevant gene signatures [57]. This may enable monitoring disease at various time-points and relate this to treatment response, efficacy as well as the possible development of resistance. However, as RNA transcription can be influenced by physiological, (pre)analytical conditions (even more than cfDNA) and clinical parameters (e.g. sex, age, inflammatory status, type and stage of disease) [11], we need to evaluate expression of each gene in healthy individuals before measurements in cancer patients become meaningful [57].

In summary, additional studies are needed to define the role of liquid biopsies at early disease stages of cancer. In particular, it must be further investigated how these biomarkers are influenced by tumor stage. This information will help to define what is the earliest stage at which a mutation can be reliably detected by the current screening methods. Then, we will benefit from these data, as they will elucidate the medical outcome of a cancer lesion facilitating the discovery of tumor that, although in early stage, will probably evolve in an unfavorable clinical course.

The role of the Estrogen Receptor in breast cancer

Breast cancer is the most common cancer and the major cause of death in women worldwide [61]. Approximately 70% of breast cancers are positive for the estrogen receptor (ER) and endocrine therapy (ET) is the mainstay of treatment for this group of patients [62]. In this scenario, the estrogen plays a crucial role in the pathogenesis and progression of breast cancer. Currently, ET include selective estrogen receptor down-regulators (e.g. fulvestrant); selective estrogen receptor modulators (e.g. tamoxifen); and aromatase inhibitors (AIs; e.g. letrozole) [7, 63, 64]. However, despite effective targeted hormonal therapies, all metastatic breast cancer patients (MBC) progress due to intrinsic endocrine resistance or the development of acquired endocrine resistance [40]. The gene coding for estrogen receptor (*ESR1*) belongs to the nuclear hormone receptor superfamily and encoded for a ligand-dependent transcription factor [65]. The receptor protein contains six domains: two-activation function (AF)-1/2 domains, a DNA binding (DBD) and hinge (H) domain and a ligand-binding domain (LBD) (**Figure 2A**). AF-2 recruits a coactivator complex consisting of one or more p160s, CREB-binding protein (CBP)/p300, and p300 and CBP-associated factor (P/CAF) [66]. The F domain attenuates 17β-estradiol-induced receptor dimerization and has a transcriptional activity [67]. In the absence of estrogen, ER is localized in the cytosol[68] but after binding of estrogen to the LBD a conformational change of

helix 12 induces the migration of the receptor from the cytosol into the nucleus, its dimerization, DNA binding and subsequent recruitment of coregulatory proteins resulting in the activation of gene transcription initiating physiological processes but also pathological processes including breast cancer tumorigenesis and progression [69]. The two *ESR1* and *ESR2* genes encode the estrogen receptor proteins ERα (66 kDa, consisting of 595 aa) and ERβ (56 kDa, consisting of 530 aa), respectively [70]. *ESR1* is located on chromosome 6 at 6q24-27 and consists of 10 exons. Beside sequence homology, the two receptors are widely expressed in normal and neoplastic breast tissue but only ERα expression has been associated to increased risk of developing breast cancer; the role or ERβ in breast cancer is less clear and no further discussed in this thesis.

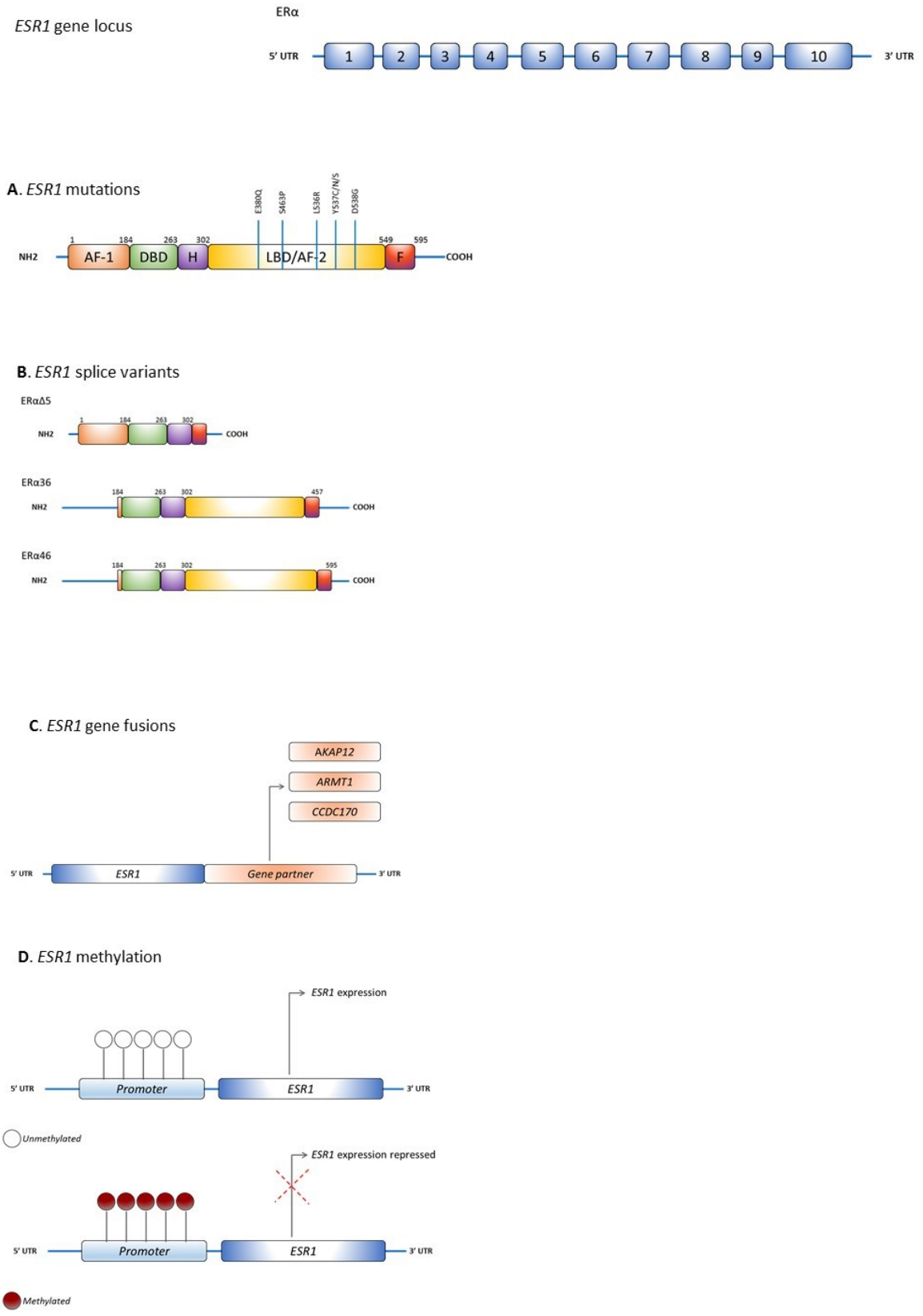


Figure 2. The *Estrogen Receptor 1 (ESR1)* gene locus and its related mechanisms of resistance

Monitoring variation in *ESR1* in endocrine resistance

To date, several ER related mechanisms has been linked to endocrine resistance, including the acquisition of activating missense mutations or fusions in *ESR1*, altered splicing of *ESR1*, and methylation of *ESR1* (**Figure 2**).

***ESR1* mutations**

Several studies have shown that up to 40% of endocrine-resistant breast cancers harbor mutations in *ESR1* [40, 48, 71, 72]. These mutations result in a constitutively active receptor that drive to a constitutive hormone-independent transcriptional activation and enhance cell proliferation. These mutations are rare or not frequently present in the primary breast tumors but are frequently acquired in metastatic tumors under the selective pressure of aromatase inhibitor (AI) therapy [63, 73, 74]. Therefore, the acquisition of *ESR1* mutations has been linked to ET resistance in MBC patients. However, patients with *ESR1* mutations can still be effectively managed by treatment with fulvestrant [46]. Although the *ESR1* mutations can occur in different sites within the ER ligand-binding domain (LBD), these mutations are most often identified at the Y537 and D538 residues of the LBD domain (**Figure 2A**) [62]. Among the different *ESR1* mutations, rare variants have not yet been functionally annotated well, while the common variants have been functionally characterized. Among the latter group, the different *ESR1* mutations either cause hormone-independent receptor activation (Y537, D538) or increased hormone sensitivity (K303R, E380Q). In addition, infrequent mutations (S432L, V534E) with no constitutive activity have been described [71, 75]. Distinct *ESR1* mutations are generally able to differentially affect the efficacy of ET therapies. Moreover, preclinical models demonstrated that cell lines transfected with ER containing the following amino acid changes D538G, Y537S, L536Q, Y537N, Y537C, S463P or E380Q are able to growth in the absence of estrogen. All this evidence suggests that identifying *ESR1* mutations provides important information to guide on treatment decision making for ER-positive MBC. In this context, the use of liquid biopsies may help to provide information on the presence of *ESR1* mutations [76].

ESR1 Splice variants

Aberrant splicing of *ESR1* has been reported as an alternative mechanism contributing to the loss of responsiveness to anti-estrogen treatment in ER-positive breast cancer (BC) patients [77,78]. *ESR1* splice variants are heterogeneously expressed in primary breast cancers and their translation result in ER protein variants with modified functions [68] which might differentially contribute to the development and progression of breast cancer. In addition to the full-length 66-kDa ERα (ERα66) which harbors the two activation functions, AF-1 and AF-2, new estrogen receptors variants have been identified (**Figure 2B**). For example, ERα30 is a novel variant of ERα in breast cancer tissue resulting from altered splicing which lacks both the LBD and the ligand-dependent transcriptional activation domain but retained the N-terminal transcriptional activation domain, the DNA-binding domain, a partial hinge domain, and possesses a unique 10-amino-acid domain [78]. Overexpression of ERα30 in MDA-MB-231 breast cancer cells caused enhanced malignant behavior [78]. ERα46 and ERα36 are 2 *ESR1* variants that can be found in normal physiologically context [68]. However, an imbalance between the two isoforms could contribute to BC [68]. ERαΔ5 is a different *ESR1* splice variant, missing the entire ligand binding domain that show a weak constitutive activity by binding DNA and interfering with ERα [79]. A study by *Beije et al*. evaluated the expression of four *ESR1* different splice variants (ERαΔ5, ERαΔ7, ERα46 and ERα36) in circulating tumor cells (CTC)-enriched mRNA. Among them, only the expression of the ERαΔ5 variant was found higher in MBC patients than in healthy blood donors despite the ERαΔ5 variant expression was not enriched during endocrine therapy for MBC [40].

ESR1 fusions

Recently, because of the advent of NGS methods, multiple *ESR1* gene fusions have been recognized in BC, but their role in disease progression and endocrine resistance remain to be understand [80]. The most common recurrent *ESR1* fusions include different genes such as *AKAP12*, *ARMT1* or *CCDC170*, which affect the breast tumorigenesis by enhanced cell-cycle progression, proliferation and survival [77]. The *ESR1* fusion transcripts are detectable in primary breast cancer, particularly in patients with high-grade disease. Structurally, fusions happen between the 5' end of *ESR1* and the 3' ends of gene partners located upstream on the same chromosome 6q25.1 (**Figure 2C**). Functionally, *ESR1*-fusions result in truncated proteins encoded

by the fusions-gene partners under an *ESR1* promoter. Therefore, the encoded chimeric proteins are generally deprived of the *ESR1* ligand-binding domain but retains both the hormone-independent transactivation domain and the DNA-binding domain [81].

ESR1 methylation

Methylation of the *ESR1* promoter has been suggested as a useful biomarker of response to therapy in breast cancer patients (**Figure 2D**). Although different studies investigated the role of *ESR1* methylation in BC, its effect on metastases and disease progression is not yet entirely clear. *ESR1* methylation is usually higher in primary and metastatic breast cancer compared to healthy breast tissue where it is virtually absent [82]. The effect of methylation was firstly described by *Lapidus et al*. showing a negative correlation between *ESR1* promoter methylation and expression of ER [83]. Moreover, several studies have demonstrated the existence of a link between the *ESR1* methylation status and a poor prognosis [84, 85]. For example, *Kirn et al*. revealed a lower survival in patients harboring *ESR1* promoter methylation in their primary breast cancer than in patients without methylation (38.1 months *vs*. 54.3 months) [82]. Interestingly, the *ESR1* promoter methylation was also associated to loss of ER expression in 37% of metastatic samples derived from this group of patients [82]. Another interesting study by *Mastoraki et al*. showed that *ESR1* methylation in CTCs correlate with paired plasma ctDNA and is strongly associated with resistance to everolimus/exemestane in patients with breast cancer [86]. All these results confirm the prognostic role of *ESR1* methylation in primary and in metastatic breast cancer.

In conclusion, this thesis is aimed at the optimization and standardization of liquid biopsies, to detect of different *ESR1* genomic alterations in breast cancer patients, and to study the relationship between (liquid) biomarkers with disease and treatment outcome. The less-invasive nature of liquid biopsies allows its use for molecular characterization of cancer lesions. By temporal follow-up, a liquid biopsy has the potential to monitor genetic changes over time, discovering acquired resistance, improving the treatment efficacy and avoid the need of repeated tissue biopsies. Therefore, liquid biopsies in my view will enable a more personalized medicine approach for the treatment of cancer patients unable to obtain by using the traditional tissue biopsies.

References

1. Biankin, A.V., S. Piantadosi, and S.J. Hollingsworth, *Patient-centric trials for therapeutic development in precision oncology.* Nature, 2015. **526**(7573): p. 361-70.
2. Marrugo-Ramirez, J., M. Mir, and J. Samitier, *Blood-Based Cancer Biomarkers in Liquid Biopsy: A Promising Non-Invasive Alternative to Tissue Biopsy.* Int J Mol Sci, 2018. **19**(10).
3. Jamal-Hanjani, M., et al., *Translational implications of tumor heterogeneity.* Clin Cancer Res, 2015. **21**(6): p. 1258-66.
4. Castro-Giner, F., et al., *Cancer Diagnosis Using a Liquid Biopsy: Challenges and Expectations.* Diagnostics (Basel), 2018. **8**(2).
5. Zhang, W., et al., *Liquid Biopsy for Cancer: Circulating Tumor Cells, Circulating Free DNA or Exosomes?* Cell Physiol Biochem, 2017. **41**(2): p. 755-768.
6. Chu, D. and B.H. Park, *Liquid biopsy: unlocking the potentials of cell-free DNA.* Virchows Arch, 2017. **471**(2): p. 147-154.
7. Guttery, D.S., et al., *Noninvasive detection of activating estrogen receptor 1 (ESR1) mutations in estrogen receptor-positive metastatic breast cancer.* Clin Chem, 2015. **61**(7): p. 974-82.
8. Mandel, P. and P. Metais, *[Not Available].* C R Seances Soc Biol Fil, 1948. **142**(3-4): p. 241-3.
9. Diaz, L.A., Jr. and A. Bardelli, *Liquid biopsies: genotyping circulating tumor DNA.* J Clin Oncol, 2014. **32**(6): p. 579-86.
10. Alix-Panabieres, C. and K. Pantel, *Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy.* Cancer Discov, 2016. **6**(5): p. 479-91.
11. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics.* Nat Med, 2008. **14**(9): p. 985-90.
12. Murtaza, M., et al., *Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA.* Nature, 2013. **497**(7447): p. 108-12.
13. Imperial, R., et al., *Matched Whole-Genome Sequencing (WGS) and Whole-Exome Sequencing (WES) of Tumor Tissue with Circulating Tumor DNA (ctDNA) Analysis: Complementary Modalities in Clinical Practice.* Cancers (Basel), 2019. **11**(9).
14. Nik-Zainal, S., et al., *Landscape of somatic mutations in 560 breast cancer whole-genome sequences.* Nature, 2016. **534**(7605): p. 47-54.

15. Holdhoff, M., et al., *Analysis of circulating tumor DNA to confirm somatic KRAS mutations.* J Natl Cancer Inst, 2009. **101**(18): p. 1284-5.
16. Diehl, F., et al., *Detection and quantification of mutations in the plasma of patients with colorectal tumors.* Proc Natl Acad Sci U S A, 2005. **102**(45): p. 16368-73.
17. Bettegowda, C., et al., *Detection of circulating tumor DNA in early- and late-stage human malignancies.* Sci Transl Med, 2014. **6**(224): p. 224ra24.
18. Deans, Z.C., et al., *IQN path ASBL report from the first European cfDNA consensus meeting: expert opinion on the minimal requirements for clinical ctDNA testing.* Virchows Arch, 2019. **474**(6): p. 681-689.
19. Abbosh, C., et al., *Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution.* Nature, 2017. **545**(7655): p. 446-451.
20. Rothwell, D.G., et al., *Genetic profiling of tumours using both circulating free DNA and circulating tumour cells isolated from the same preserved whole blood sample.* Mol Oncol, 2016. **10**(4): p. 566-74.
21. van Dessel, L.F., et al., *Application of circulating tumor DNA in prospective clinical oncology trials - standardization of preanalytical conditions.* Mol Oncol, 2017. **11**(3): p. 295-304.
22. Laulagnier, K., et al., *Mast cell- and dendritic cell-derived exosomes display a specific lipid composition and an unusual membrane organization.* Biochem J, 2004. **380**(Pt 1): p. 161-71.
23. Zaborowski, M.P., et al., *Extracellular Vesicles: Composition, Biological Relevance, and Methods of Study.* Bioscience, 2015. **65**(8): p. 783-797.
24. Valadi, H., et al., *Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells.* Nat Cell Biol, 2007. **9**(6): p. 654-9.
25. Heijnen, H.F., et al., *Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules.* Blood, 1999. **94**(11): p. 3791-9.
26. Cocucci, E. and J. Meldolesi, *Ectosomes and exosomes: shedding the confusion between extracellular vesicles.* Trends Cell Biol, 2015. **25**(6): p. 364-72.
27. S, E.L.A., et al., *Extracellular vesicles: biology and emerging therapeutic opportunities.* Nat Rev Drug Discov, 2013. **12**(5): p. 347-57.

28. Thery, C., et al., *Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles.* J Immunol, 2001. **166**(12): p. 7309-18.
29. Skog, J., et al., *Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers.* Nat Cell Biol, 2008. **10**(12): p. 1470-6.
30. Svensson, K.J., et al., *Hypoxia triggers a proangiogenic pathway involving cancer cell microvesicles and PAR-2-mediated heparin-binding EGF signaling in endothelial cells.* Proc Natl Acad Sci U S A, 2011. **108**(32): p. 13147-52.
31. Grange, C., et al., *Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche.* Cancer Res, 2011. **71**(15): p. 5346-56.
32. Millimaggi, D., et al., *Tumor vesicle-associated CD147 modulates the angiogenic capability of endothelial cells.* Neoplasia, 2007. **9**(4): p. 349-57.
33. Antonyak, M.A., et al., *Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fibronectin to recipient cells.* Proc Natl Acad Sci U S A, 2011. **108**(12): p. 4852-7.
34. Balaj, L., et al., *Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences.* Nat Commun, 2011. **2**: p. 180.
35. O'Brien, K., et al., *RNA delivery by extracellular vesicles in mammalian cells and its applications.* Nat Rev Mol Cell Biol, 2020.
36. Heitzer, E., P. Ulz, and J.B. Geigl, *Circulating tumor DNA as a liquid biopsy for cancer.* Clin Chem, 2015. **61**(1): p. 112-23.
37. Tong, Y., et al., *Application of Digital PCR in Detecting Human Diseases Associated Gene Mutation.* Cell Physiol Biochem, 2017. **43**(4): p. 1718-1730.
38. Postel, M., et al., *Droplet-based digital PCR and next generation sequencing for monitoring circulating tumor DNA: a cancer diagnostic perspective.* Expert Rev Mol Diagn, 2018. **18**(1): p. 7-17.
39. Elazezy, M. and S.A. Joosse, *Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management.* Comput Struct Biotechnol J, 2018. **16**: p. 370-378.
40. Beije, N., et al., *Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients.* Mol Oncol, 2018. **12**(1): p. 48-57.

41. Del Re, M., et al., *The Detection of Androgen Receptor Splice Variant 7 in Plasma-derived Exosomal RNA Strongly Predicts Resistance to Hormonal Therapy in Metastatic Prostate Cancer Patients.* Eur Urol, 2017. **71**(4): p. 680-687.
42. Garcia, J., et al., *Evaluation of pre-analytical conditions and comparison of the performance of several digital PCR assays for the detection of major EGFR mutations in circulating DNA from non-small cell lung cancers: the CIRCAN_0 study.* Oncotarget, 2017. **8**(50): p. 87980-87996.
43. Olmedillas-Lopez, S., M. Garcia-Arranz, and D. Garcia-Olmo, *Current and Emerging Applications of Droplet Digital PCR in Oncology.* Mol Diagn Ther, 2017. **21**(5): p. 493-510.
44. Riediger, A.L., et al., *Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients.* Sci Rep, 2016. **6**: p. 33505.
45. Misale, S., et al., *Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer.* Nature, 2012. **486**(7404): p. 532-6.
46. Fribbens, C., et al., *Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer.* J Clin Oncol, 2016. **34**(25): p. 2961-8.
47. Takeshita, T., et al., *Clinical significance of monitoring ESR1 mutations in circulating cell-free DNA in estrogen receptor positive breast cancer patients.* Oncotarget, 2016. **7**(22): p. 32504-18.
48. Wang, P., et al., *Sensitive Detection of Mono- and Polyclonal ESR1 Mutations in Primary Tumors, Metastatic Lesions, and Cell-Free DNA of Breast Cancer Patients.* Clin Cancer Res, 2016. **22**(5): p. 1130-7.
49. Takeshita, T., et al., *Prognostic role of PIK3CA mutations of cell-free DNA in early-stage triple negative breast cancer.* Cancer Sci, 2015. **106**(11): p. 1582-9.
50. Ferracin, M., et al., *Absolute quantification of cell-free microRNAs in cancer patients.* Oncotarget, 2015. **6**(16): p. 14545-55.
51. Chen, W.W., et al., *BEAMing and Droplet Digital PCR Analysis of Mutant IDH1 mRNA in Glioma Patient Serum and Cerebrospinal Fluid Extracellular Vesicles.* Mol Ther Nucleic Acids, 2013. **2**: p. e109.
52. Stahlberg, A., et al., *Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing.* Nucleic Acids Res, 2016. **44**(11): p. e105.

53. Pietrasz, D., et al., *Plasma Circulating Tumor DNA in Pancreatic Cancer Patients Is a Prognostic Marker.* Clin Cancer Res, 2017. **23**(1): p. 116-123.
54. Chae, Y.K., et al., *Concordance of Genomic Alterations by Next-Generation Sequencing in Tumor Tissue versus Circulating Tumor DNA in Breast Cancer.* Mol Cancer Ther, 2017. **16**(7): p. 1412-1420.
55. Heitzer, E., et al., *Current and future perspectives of liquid biopsies in genomics-driven oncology.* Nat Rev Genet, 2019. **20**(2): p. 71-88.
56. Heitzer, E., et al., *The potential of liquid biopsies for the early detection of cancer.* NPJ Precis Oncol, 2017. **1**(1): p. 36.
57. Perakis, S. and M.R. Speicher, *Emerging concepts in liquid biopsies.* BMC Med, 2017. **15**(1): p. 75.
58. Tie, J., et al., *Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer.* Sci Transl Med, 2016. **8**(346): p. 346ra92.
59. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
60. Narasimhan, V.M., et al., *Health and population effects of rare gene knockouts in adult humans with related parents.* Science, 2016. **352**(6284): p. 474-7.
61. Li, T., et al., *A Meta-Analysis of the Association between ESR1 Genetic Variants and the Risk of Breast Cancer.* PLoS One, 2016. **11**(4): p. e0153314.
62. Dustin, D., G. Gu, and S.A.W. Fuqua, *ESR1 mutations in breast cancer.* Cancer, 2019. **125**(21): p. 3714-3728.
63. Robinson, D.R., et al., *Activating ESR1 mutations in hormone-resistant metastatic breast cancer.* Nat Genet, 2013. **45**(12): p. 1446-51.
64. Riggins, R.B., et al., *Pathways to tamoxifen resistance.* Cancer Lett, 2007. **256**(1): p. 1-24.
65. Olefsky, J.M., *Nuclear receptor minireview series.* J Biol Chem, 2001. **276**(40): p. 36863-4.
66. Webb, P., et al., *Estrogen receptor activation function 1 works by binding p160 coactivator proteins.* Mol Endocrinol, 1998. **12**(10): p. 1605-18.
67. Yang, J., et al., *The F-domain of estrogen receptor-alpha inhibits ligand induced receptor dimerization.* Mol Cell Endocrinol, 2008. **295**(1-2): p. 94-100.
68. Inoue, K. and E.A. Fry, *Aberrant Splicing of Estrogen Receptor, HER2, and CD44 Genes in Breast Cancer.* Genet Epigenet, 2015. **7**: p. 19-32.
69. Heldring, N., et al., *Estrogen receptors: how do they signal and what are their targets.* Physiol Rev, 2007. **87**(3): p. 905-31.

70. Thomas, C. and J.A. Gustafsson, *Estrogen receptor mutations and functional consequences for breast cancer.* Trends Endocrinol Metab, 2015. **26**(9): p. 467-76.
71. Chandarlapaty, S., et al., *Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial.* JAMA Oncol, 2016. **2**(10): p. 1310-1315.
72. Schiavon, G., et al., *Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer.* Sci Transl Med, 2015. **7**(313): p. 313ra182.
73. Jeselsohn, R., et al., *Allele-Specific Chromatin Recruitment and Therapeutic Vulnerabilities of ESR1 Activating Mutations.* Cancer Cell, 2018. **33**(2): p. 173-186 e5.
74. Toy, W., et al., *ESR1 ligand-binding domain mutations in hormone-resistant breast cancer.* Nat Genet, 2013. **45**(12): p. 1439-45.
75. Toy, W., et al., *Activating ESR1 Mutations Differentially Affect the Efficacy of ER Antagonists.* Cancer Discov, 2017. **7**(3): p. 277-287.
76. Angus, L., et al., *ESR1 mutations: Moving towards guiding treatment decision-making in metastatic breast cancer patients.* Cancer Treat Rev, 2017. **52**: p. 33-40.
77. Li, G., et al., *Estrogen receptor-alpha36 is involved in development of acquired tamoxifen resistance via regulating the growth status switch in breast cancer cells.* Mol Oncol, 2013. **7**(3): p. 611-24.
78. Zhu, H., et al., *Identification of a novel human estrogen receptor-alpha splice variant able to enhance malignant biological behaviors of breast cancer cells.* Oncol Lett, 2018. **15**(4): p. 5339-5344.
79. Taylor, S.E., et al., *Elevated oestrogen receptor splice variant ERalphaDelta5 expression in tumour-adjacent hormone-responsive tissue.* Int J Environ Res Public Health, 2010. **7**(11): p. 3871-89.
80. Giltnane, J.M., et al., *Genomic profiling of ER(+) breast cancers after short-term estrogen suppression reveals alterations associated with endocrine resistance.* Sci Transl Med, 2017. **9**(402).
81. Veeraraghavan, J., et al., *Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers.* Nat Commun, 2014. **5**: p. 4577.
82. Kirn, V., et al., *ESR1-promoter-methylation status in primary breast cancer and its corresponding metastases.* Clin Exp Metastasis, 2018. **35**(7): p. 707-712.

83. Lapidus, R.G., et al., *Methylation of estrogen and progesterone receptor gene 5' CpG islands correlates with lack of estrogen and progesterone receptor gene expression in breast tumors.* Clin Cancer Res, 1996. **2**(5): p. 805-10.

84. Muller, H.M., et al., *DNA methylation in serum of breast cancer patients: an independent prognostic marker.* Cancer Res, 2003. **63**(22): p. 7641-5.

85. Widschwendter, M., et al., *Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen.* Cancer Res, 2004. **64**(11): p. 3807-13.

86. Mastoraki, S., et al., *ESR1 Methylation: A Liquid Biopsy-Based Epigenetic Assay for the Follow-up of Patients with Metastatic Breast Cancer Receiving Endocrine Treatment.* Clin Cancer Res, 2018. **24**(6): p. 1500-1510.

Chapter 2

Aims and outline of this thesis

Aims and outline of this thesis

The overall aim of this thesis was to address several technical and clinical aspects of liquid biopsies and to determine the clinical relevance of *ESR1* genomic alterations in breast cancer. By combining and evaluating different circulating biomarkers, we aimed to establish the optimal pre-analytical conditions and to improve the liquid biopsies used in standard clinical care. Furthermore, we defined proper methods to investigate rare mutations in circulating biomarkers and accurately assess its relevance with regard to the tumor's clinical behavior. These analyses combined will enable us to better investigate the relation between liquid biopsies and treatment outcome of the patient's malignant lesion ultimately contributing to more personalized healthcare.

To achieve all this, I investigated the various relevant pre-analytical, analytical and post-analytical conditions required for successful routine diagnostics using liquid biopsies as starting material. For high-throughput processing of plasma samples from large prospective multi-center studies, we assessed whether automated platforms perform equally well, with respect to quality/quantity of ctDNA isolated, as compared to the current 'gold standard' manual system (**Chapter 3**)**.** The validity of using ctDNA as a clinical biomarker, compared to DNA isolated from tissue biopsy, was further extended by measuring somatic variants of *TP53* in patients with high-grade serous ovarian cancer (**Chapter 4**). Next, I evaluated biofluid-derived RNA, which might add clinical value to assess tumor-specific changes and allow the identification of tumor-specific gene expression profiles. To this purpose, a workflow for the isolation and characterization of EV-RNA from cancer patients was set up and evaluated for the detection of mutant transcripts compared to simultaneously collected ctDNA (**Chapter 5**). To analyze ctDNA in minute amounts of plasma from retrospective studies, we set up a workflow to detect *ESR1* mutations, by both dPCR and NGS, in metastatic breast cancer patients (**Chapter 6**). Finally, with the aim to transfer this information in the future to liquid biopsies, we investigated the prognostic and predictive value of three different fusion genes (*ESR1* with *CCDC170*, *AKAP12* and *C6orf211/ARMT1,* respectively*)* in breast cancer patients with early disease or with advanced disease treated with endocrine therapy (**Chapter 7**)**.** Taken together the overarching aim of **this thesis** was to develop and standardize methods and to evaluate whether liquid biopsies and/or *ESR1* genomic alterations are suitable as biomarkers for clinical disease monitoring as well as for the assessment of treatment outcome.

Chapter 3

High-throughput isolation of circulating tumor DNA: a comparison of automated platforms

Vitale SR*, van Dessel LF*, Helmijr JCA, Wilting SM, van der Vlugt-Daane M, Oomen-de Hoop E, Sleijfer S, Martens JWM, Jansen MPHM, Lolkema MP

Mol Oncol. 2019 Feb; 13(2):392-402
doi: 10.1002/1878-0261.12415

Abstract

The emerging interest in circulating tumor DNA (ctDNA) analyses for clinical trials has necessitated the development of a high-throughput method for fast, reproducible, and efficient isolation of ctDNA. Currently, the majority of ctDNA studies use the manual QIAamp (QA) platform to isolate DNA from blood. The purpose of this study was to compare two competing automated DNA isolation platforms [Maxwell (MX) and QIAsymphony (QS)] to the current 'gold standard' QA to facilitate high-throughput processing of samples in prospective trials.

We obtained blood samples from healthy blood donors and metastatic cancer patients for plasma isolation. Total cell-free DNA (cfDNA) quantity was assessed by TERT quantitative PCR. Recovery efficiency was investigated by quantitative PCR analysis of spiked-in synthetic plant DNA. In addition, a b-actin fragmentation assay was performed to determine the amount of contamination by genomic DNA from lysed leukocytes. ctDNA quality was assessed by digital PCR for somatic variant detection. cfDNA quantity and recovery efficiency were lowest using the MX platform, whereas QA and QS showed a comparable performance.

All platforms preferentially isolated small (136 bp) DNA fragments over large (420 and 2000 bp) DNA fragments. Detection of the number variant and wild-type molecules was most comparable between QA and QS. However, there was no significant difference in variant allele frequency comparing QS and MX to QA.

In summary, we show that the QS platform has comparable performance to QA, the 'gold standard', and outperformed the MX platform depending on the readout used. We conclude that the QS can replace the more laborious QA platform, especially when high-throughput cfDNA isolation is needed.

Introduction

With the discovery of cell-free DNA (cfDNA), first described in 1948 by Mandel and Metais [1], and subsequently circulating tumor DNA (ctDNA) [2], a novel biomarker in cancer research became available. Since then, many studies have shown its great potential for detecting minimal residual disease and evaluating treatment response [3-11]. However, to enable high throughput ctDNA analyses a fast, accurate, and efficient cfDNA isolation method is highly needed.

Currently, the majority of ctDNA studies use Qiagen's QIAamp (QA) platform for cfDNA isolation [12-14]. However, this manual platform is laborious and can only process up to 24 samples at a time rendering this method less suitable for large-scale studies. Automation of cfDNA isolation represents a potential solution provided that it is able to (a) reduce hands-on time; (b) simultaneously process large numbers of samples; (c) accurately and reproducibly isolate cfDNA with a reasonable recovery; and (d) preserve the quality of ctDNA for downstream analyses.
Cell-free DNA is naturally fragmented (140–175 bp) and only present at low concentrations in the blood circulation (usually around 10 ng per mL plasma) [15]. In addition, the fraction of ctDNA relative to cfDNA can vary from extremely low (< 0.01%) to very high (60%), as it is dependent on tumor type and stage [6, 16]. Together these features make it imperative to carefully determine the efficacy of DNA isolation instead of merely investigating isolation yields. Furthermore, isolation of cfDNA and ctDNA therein is highly susceptible to genomic DNA contamination from lysed leukocytes [17, 18], resulting in a potential underestimation of the ctDNA fraction and decreasing the detection sensitivity. As potential differences in cfDNA recovery efficiency between isolation methods might affect downstream analysis results of ctDNA by decreasing its detection sensitivity, standardized comparison of the different methods for cfDNA isolation is important and highly needed.

The purpose of this study was to compare two automated cfDNA isolation platforms, Maxwell (MX) and QIAsymphony (QS), to the current 'gold standard' QA isolation kit to determine whether these automated platforms can facilitate high-throughput processing of samples in prospective trials. Our analyses focused on both qualitative and quantitative parameters, including cfDNA yield, recovery efficiency, cfDNA fragmentation patterns, and ctDNA fraction retrieved, using optimally processed plasma samples of healthy blood donors (HBDs) and patients with metastatic cancer.

Materials and methods

Subjects

Blood samples were obtained from a total of 10 HBDs and 10 metastatic cancer patients. HBDs were either laboratory volunteers or blood donors of the Sanquin Blood Bank South-West Region, The Netherlands. Patients were enrolled in this study between September 2016 and September 2017 within the Erasmus MC Cancer Institute in Rotterdam, the Netherlands. Eligibility criteria for patients have been described previously [19]. All patients provided written informed consent, and the institutional review board approved the protocols (Erasmus MC ID MEC 15-616). The study methodologies conformed to the standards set by the Declaration of Helsinki. Patient and tumor characteristics are summarized in **Table 1**.

Table 1. Patient and tumor characteristics.

Patient ID (#)	Primary tumor	Known somatic variant (nucleotide change)	Variant allele frequency in tissue (%)
BP-001	NSCLC	KRAS p.G12C (c.34G>T)	32
BP-003	Melanoma	NRAS p.Q61R (c.182A>G)	88
BP-004	Melanoma	BRAF p.V600E (c.1799_1800delinsAA)	50
BP-007	Melanoma	BRAF p.V600K (c.1798_1799delGTinsAA)	38
BP-008	CRC	KRAS p.G12D (c.35G>A)	45
BP-009	CRC	PIK3CA p.E545K (c.1633G>A)	45
BP-015	CRC	KRAS p.G13D (c.38G>A)	40
BP-016	CRC	KRAS p.G12V (c.35G>T)	Unknown
BP-023	CRC	KRAS p.G13D (c.38G>A)	Unknown
BP-028	Melanoma	BRAF p.V600K (c.1798_1799delinsAA)	55

CRC: colorectal cancer; NSCLC: non-small cell lung cancer.

Blood collection

Healthy blood donors donated 20 mL of blood, collected either in 2 x 10 mL CellSave preservative tubes (Janssen Diagnostics, Raritan, NJ, USA) or in 1 x 10 mL EDTA tube (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) and 1 x 10 mL Cell- Save preservative tube. Patients donated 3 x 10 mL of blood collected in CellSave preservative tubes. Blood samples were stored at room temperature until further processing. After blood draw, samples in EDTA tubes were processed

within 24 h, whereas samples in Cell- Save tubes were processed within 96 h for plasma isolation as previously described [19].

cfDNA isolation

Cell-free DNA was isolated from 2 mL of plasma and eluted in 60 µL of the provided elution buffer. Three isolation platforms were evaluated (**Table 2**):

- QIAamp(QA) Circulating Nucleic Acid Kit (Qiagen, Hilden, North Rhine-Westphalia, Germany);
- QIAsymphony(QS) SP Circulating DNA Kit (Qiagen);
- Maxwell(MX) RSC LV ccfDNA Plasma Custom Kit (Promega, Madison, WI, USA).

Table 2. Specifications of cell-free DNA isolation platforms.

Platform	Manufacturer	Protocol	cfDNA isolation kit	Plasma input (mL)	Number of samples per run	Handling-time per run (min)	Technique	Cost(€) per sample
QIAamp (QA)	Qiagen	Manual	QIAamp® Circulating Nucleic Acid Kit	1.0 – 5.0	24	180-240	Vacuum-column based	20
QIAsymphony (QS)	Qiagen	Automatic	QIAsymphony® Circulating DNA Kit	2.0 – 8.0*	96	30	Magnetic-bead based	24
Maxwell (MX)	Promega	Automatic	Maxwell® RSC LV ccfDNA Plasma Custom Kit	2.0 – 4.0*	16**	30	Magnetic-bead based	20

*Upon request the manufacturer is able to adjust system settings and protocols for lower/higher plasma input volumes.
** The Maxwell RSC 48 Instrument can process up to 48 samples per run.

All cfDNA isolations were performed according to the manufacturer's protocol, with some minor modifications. In more detail, cfDNA was isolated with QA as previously described [19]. The QS isolation was adapted by adding 1 µg of carrier RNA (cRNA, Qiagen) to the plasma sample preceding isolation. Using the MX platform, a third plasma centrifugation step at 2000 g for 10 min at room temperature was performed after thawing to eliminate residual leukocytes, as recommended by the manufacturer. The custom Maxwell l® RSC ccfDNA Plasma Kit for large

plasma volume protocol was used. In brief, 2 mL of plasma was added to an equal amount of binding buffer and 140 µL of magnetic beads. This mixture was incubated under rotation for 45 min at room temperature and subsequently centrifuged at 2000 g for 1 min at room temperature. The pelleted mix of beads and cfDNA was then transferred to the cartridge and run on the MX instrument (Promega) according to the manufacturer's protocol.

Testing of cRNA addition to the automated platforms

Plasma samples from several HBDs were pooled and divided into aliquots of 2 mL each. To each aliquot, we added different amounts of cRNA, ranging from 0.25 up to 4 µg. As a control, plasma samples without cRNA were included. To allow determination of the recovery efficiency, synthetic plant DNA was added to plasma samples (see below).

Cell-free DNA quantification

All cfDNA samples were quantified by both Qubit™ fluorometric quantitation (Invitrogen, Life Technologies, Carlsbad, CA, USA) and human TaqMan® copy number reference assay TERT (Applied Biosystems, Life Technologies, Foster City, CA, USA) by quantitative PCR (qPCR). The Qubit™ measurement was performed on 2 µL of each cfDNA sample using the Quant-iT dsDNA high-sensitivity assay (Invitrogen), according to the manufacturer's protocol. TERT qPCRs contained 5 µL cfDNA, 3.13 µL SensiFAST™ SYBR® Lo-Rox mix (Bioline, London, UK), and 0.62 µL TERT assay in a total reaction volume of 12.5 µL. The qPCR was performed on an Mx3000P Real-Time PCR System (Agilent, Santa Clara, CA, USA) with a pre-incubation at 95 °C for 10 min, followed by 45 cycles of 95 °C for 10 s and 60 °C for 22 s. cfDNA was quantified using a standard curve of human genomic DNA.

Synthetic plant DNA and plant DNA qPCR assay

The synthetic plant DNA assay developed by Kang et al. [20] was used as an exogenous control to calculate the recovery efficiency of each cfDNA isolation method. In short, 250 ng of a 150 bp gBlocks® gene fragment (Integrated DNA Technologies Incorporation (IDT), Coralville, Iowa, USA) was re-suspended in LoTE buffer to a final concentration of $1.64x10^{0}$ ng/µL^{-1}. The stock sample was serially diluted to a final concentration of $1.64x10^{-6}$ ng/µL^{-1} of which 5 µL was spiked into

plasma preceding cfDNA isolation. Plant DNA qPCR reactions were essentially performed as described above, using 900 nM of both forward and reverse primer and 250 nM of a FAM-labeled probe (**Table S1**). Recovery efficiency was determined using a standard curve including the amount of spiked-in plant DNA. Samples with a recovery efficiency <5% or >100% were excluded from further analysis as this strongly suggested an operator failure. This was further supported by the fact that recovery efficiency was not strongly correlated (ρ=0.45) with cfDNA concentration (**Figure S1**).

Digital PCR TaqMan SNP genotyping and b-actin fragmentation assay

The presence of somatic tumor-specific variants and wild-type DNA molecules was determined using standard and custom-made TaqMan Single Nucleotide Polymorphism (SNP) genotyping assays (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's instructions (Tables S2 and S3). The TaqMan β-actin assay was used to investigate the fragment size distribution as an indication of leukocyte DNA contamination of the cfDNA, as previously reported [19]. In short, a standard amount of 2 ng of cfDNA was used to detect one small (136 bp) and two long (420 and 2000 bp) b-actin fragments within a single reaction. The used primers and probes are indicated in Table S1. The digital PCR (dPCR) was performed as previously described [19]. In short, a maximum volume input of 7.8 µL of the final cfDNA eluate was added to the dPCR; the dPCR run was performed on the chip-based QuantStudio 3D Digital PCR System (Thermo Fisher Scientific) according to the manufacturer's protocol. SNP genotyping assays were run at 56 °C; the b-actin assay was run at 60 °C. A negative control (H2O) and a positive control (cell genomic DNA with known variant) were added to every experiment.

Sample size

To test whether QS and MX were comparable to QA, we assumed a Cohen's effect size of 0.8, to be able to detect relevant differences. With a two-sided type I error probability (a) of 0.025 and a type II error probability (b) of 0.2, a power calculation determined that 18 subjects were needed for paired comparisons. Based on the foregoing, 20 subjects were included (10 HBDs and 10 patients).

Calculations and statistical analysis

All assay results were corrected for variations in plasma input and eluate volume, as previously described [19] and expressed as either ng/mL plasma or as mutant/wild type/β-actin copy number per mL of plasma.

The variant allele frequency (VAF) was calculated as follows: VAF = total variant copy number/(total variant copy number + total wild-type copy number). The statistical analyses and figure plotting were performed in R version 3.2.3. The Friedman test was used to test the difference between matched QA, MX, and QS samples. Significant differences were post hoc analysed using the Wilcoxon signed-rank test. To correct for multiple testing, we adjusted the P value for significance by subsequently applying the Bonferroni correction. The Wilcoxon signed-rank test was used to test the difference between matched EDTA and CellSave samples. Correlations were determined by Spearman's rank correlation coefficient.

Results

Optimization of cfDNA isolation using automated isolation platforms

In a small pilot study, we had previously observed a beneficial effect of cRNA addition to HBD plasma during isolation with the QS protocol on the cfDNA yield as determined by Qubit (**Figure S2**). Therefore, cRNA addition was implemented in our standard QS protocol. However, it has been reported that cRNA might interfere with Qubit-based DNA quantification and might not be a reliable readout [19]. Therefore, we tested whether cfDNA isolation on the automated platforms (QS/MX) was beneficially or adversely affected by the addition of cRNA using multiple readouts. We added varying amounts of cRNA to the plasma samples and measured the resulting cfDNA concentration by Qubit and TERT qPCR for both automated platforms. Using Qubit as readout, the addition of cRNA increased the total amount of cfDNA extracted on both platforms (MX $P < 0.001$; QS $P < 0.001$; **Figure 1A**). However, using TERT qPCR as readout, this increase could not be reproduced (**Figure 1B**). Next, we assessed the impact of cRNA on the recovery of spiked-in synthetic plant DNA. Addition of cRNA affected the recovery efficiency of plant DNA (MX $P = 0.02$; QS $P = 0.04$; **Figure 1C**). Independent of cRNA input, recovery of plant DNA was ~ 30% higher with QS (58.37 9.52) than with MX (28.22 6.67; $P < 0.001$). To assess whether the addition of cRNA biased the isolation of particular cfDNA fragment sizes, we performed the b-

actin fragmentation assay (**Figure 1D**). For both methods, increasing amounts of cRNA reduced the number of small fragments (136 bp; MX P = 0.001; QS P < 0.001), while no effect on larger fragments was observed. For all post hoc analyses, paired testing of samples with and without addition of cRNA (0 µg) did not reveal any significant differences.

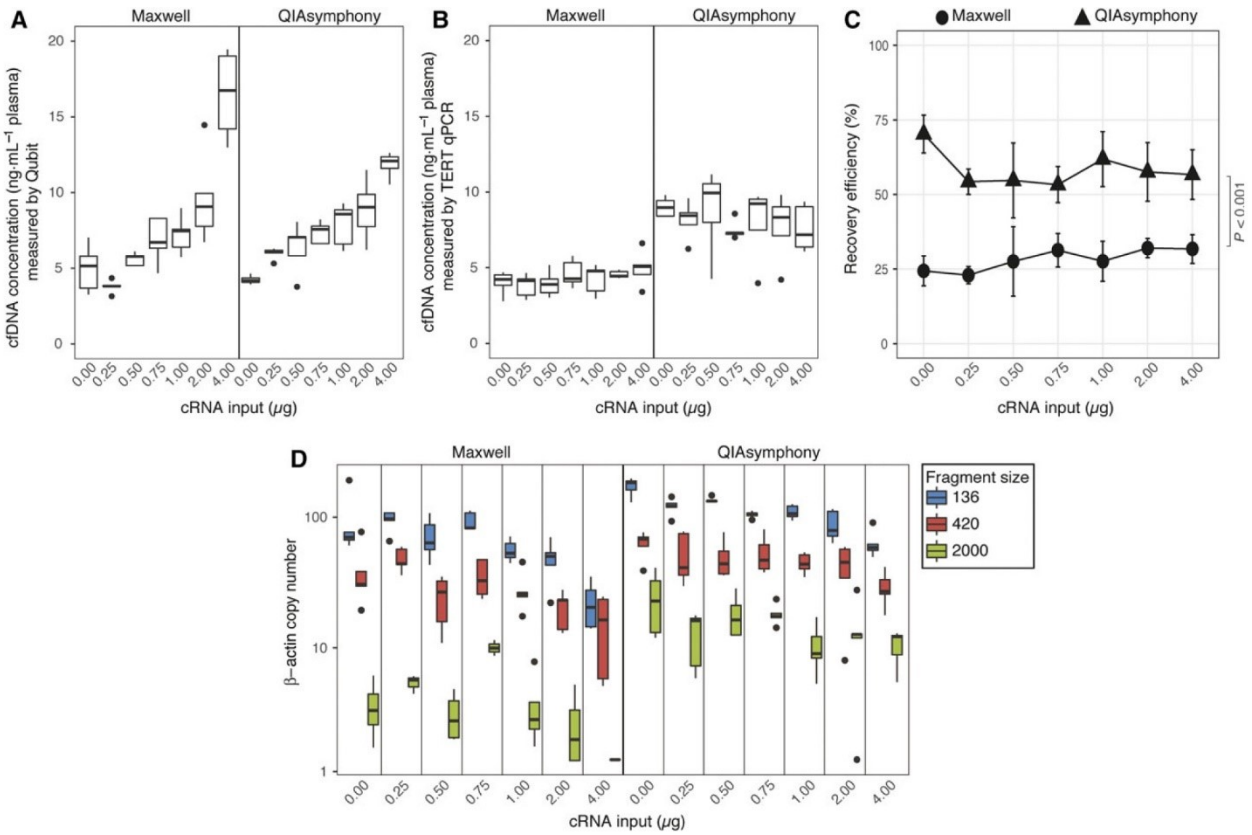


Figure 1. Effect of increasing cRNA input (0-4 µg) on cfDNA quantity and quality using the Maxwell and QIAsymphony platforms.

The effect on cfDNA concentration (ng/mL plasma) was measured by Qubit (A) and TERT qPCR (B). The recovery efficiency of each platform was analyzed by qPCR using spiked-in synthetic plant DNA (C). Differences in cfDNA fragment size, expressed as number of β-actin fragments for each fragment size (136 bp, 420 bp and 2000 bp), was analyzed by dPCR (D). Boxes (interquartile ranges; IQR) and whiskers (1.5x IQR) are shown together with the median (black horizontal line). Outliers are indicated as single black points. Symbols ● and ▲are mean values shown with whiskers (standard deviation). The Friedman test was used to test the group difference between Maxwell and QIAsymphony samples. Significant differences were post-hoc analyzed using the Wilcoxon signed rank test. N = 5.

Compatibility of CellSave preservative tubes with different isolation platforms

Previously, we have demonstrated the good performance of CellSave preservative tubes for ctDNA analysis [19]. However, the manufacturers of both automated platforms recommend to use plasma isolated from blood collected in EDTA tubes. To allow for a fair comparison with our CellSave QA results, we therefore first determined whether the automated platforms (QS/MX) were compatible with CellSave tubes by assessing the cfDNA quantity and quality.

Figure 2A shows cfDNA concentrations as measured by TERT qPCR analysis. For the MX platform, the median cfDNA concentration was 5.59 ng x mL $^{-1}$ plasma from EDTA tubes and was 2.19 ng x mL $^{-1}$ plasma from CellSave tubes (IQR: 5.06–6.21 and 2.07– 3.37 ng x mL $^{-1}$ plasma, respectively; P = 0.008). For the QS platform, the median cfDNA concentration was 17.17 ng x mL $^{-1}$ plasma from EDTA tubes and 11.13 ng ng x mL $^{-1}$ plasma from CellSave tubes (IQR: 7.81– 22.12 and 9.02–14.14 ng x mL $^{-1}$ plasma, respectively).

Although this was comparable, EDTA samples displayed a larger range in yielded cfDNA concentration. The potential effect of CellSave tubes on the recovery of synthetic plant DNA was determined as well. Comparable recovery efficiencies were observed in plasma collected in EDTA and CellSave tubes for both platforms (39.92% vs. 44.27% in MX and 67.92% vs. 66.19% in QS; **Figure 2B**). Finally, we used the b-actin fragmentation assay to evaluate cfDNA fragmentation patterns as a readout for general sample quality (**Figure 2C**). EDTA tubes yielded a higher number of large cfDNA fragments (2000 bp) irrespective of the platform used (median number of b-actin fragments and IQR MX: 33.08 (14.28–44.59); QS: 32.46 (25.53– 55.44)) than CellSave tubes (median number of b-actin fragments and IQR MX: 5.15 (2.42–9.17); QS: 13.80 (7.01–18.18); P = 0.008). The number of small DNA fragments (136 bp) did not differ between EDTA and CellSave tubes for MX, but was slightly higher for EDTA tubes on the QS platform (median number of bactin fragments and IQR EDTA: 142.71 (110.28– 198.18); CellSave: 89.71 (80.22–102.64); P = 0.04). Based on these results, we deemed CellSave tubes are compatible with both automated platforms and used them for all further experiments.

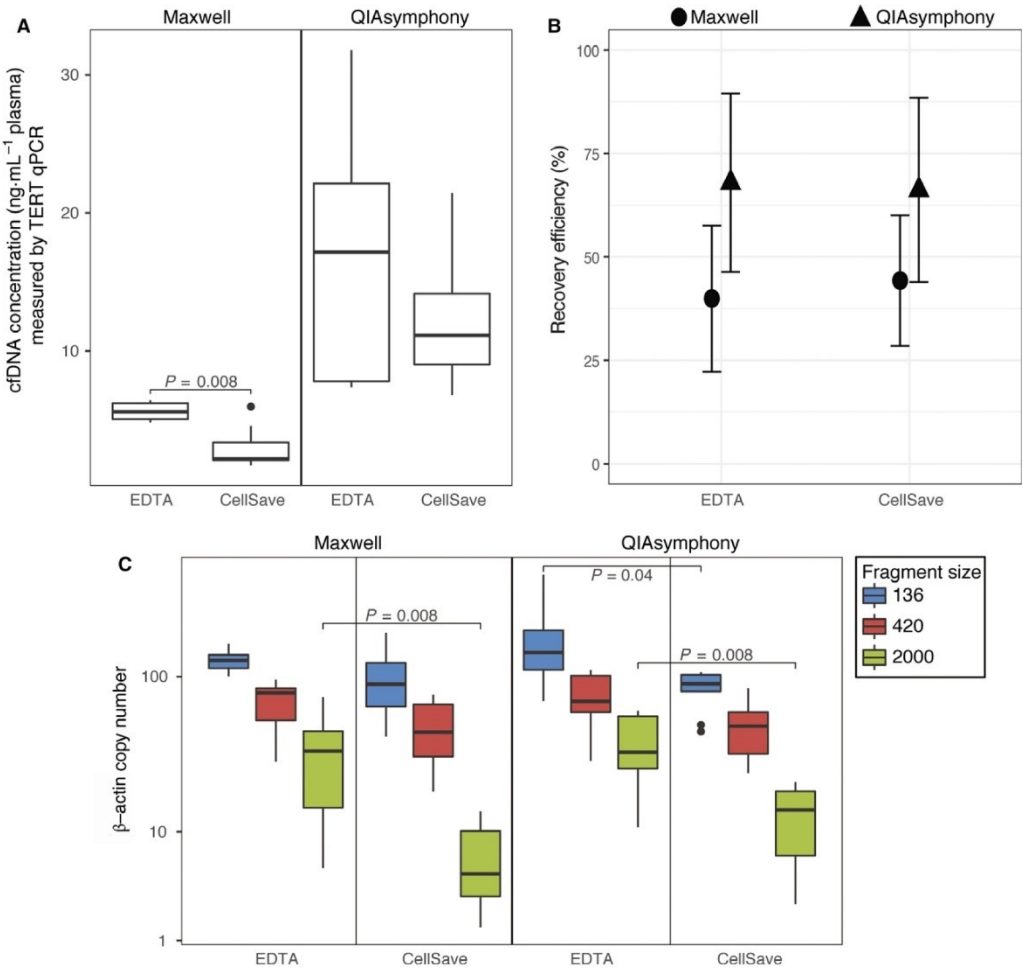


Figure 2. Compatibility of EDTA and CellSave blood collection tubes with the Maxwell and QIAsymphony platforms.

The effect on cfDNA concentration (ng/mL plasma) measured by TERT qPCR, (A) recovery efficiency measured by plant DNA qPCR and (B) β-actin fragmentation assay (C) analyzed with dPCR are shown. Boxes (interquartile ranges; IQR) and whiskers (1.5x IQR) are shown together with the median (black horizontal line). Outliers are indicated as single black points. Symbols ● and ▲are mean values shown with whiskers (standard deviation). The Wilcoxon signed rank test was used to test the difference between blood collection tubes for each platform. N = 9.

Comparison of the performance of automated platforms on downstream cfDNA and ctDNA analyses

Next, we compared the quantity and quality of the obtained cfDNA using the current 'gold standard' manual QA platform to the automated QS and MX platforms using samples from 10 HBDs and 10 metastatic cancer patients. In HBDs, cfDNA concentrations measured by TERT qPCR analysis were comparable for all three isolation platforms (**Figure 3A**). In patients, the MX retrieved significantly less cfDNA compared to both QA (P = 0.002) and QS (P = 0.002; median cfDNA concentration and IQR QA: 15.84 (12.64–65.11); MX: 6.00 (3.80–20.43); QS: 14.50 (11.99–57.65) ng x mL $^{-1}$ plasma; **Figure 3A**). To determine the recovery efficiency of the three different platforms, 5 µL of synthetic plant DNA was added to each plasma sample preceding cfDNA isolation. The average recovery efficiency using QA (51.95 12.02%) was similar to QS (43.45 8.21%). However, MX performed worse (18.61 5.81%; P < 0.001; Fig. 3B). In HBDs, we did not observe cfDNA fragment size differences between either of the evaluated platforms (**Figure 3C**). In patients, MX isolated fewer small b-actin fragments (136 bp) than QA (median number of b-actin fragments and IQR for MX: 57.45 (53.17–66.72); and for QA: 83.18 (70.36–101.63); P < 0.01) and fewer large fragments (2000 bp) than QS (median number of bactin fragments and IQR for MX: 2.08 (0.00–5.21); and for QS: 10.06 (6.70–13.72); P = 0.002). Finally, we compared somatic variant detection in ctDNA isolated by the different platforms. For this purpose, we used previously generated diagnostic sequencing results on the somatic variant status in the primary and/or metastatic lesions of the corresponding patients (**Table 1)**. We detected the expected somatic variants in all patients for all isolation methods. QS results were most comparable to QA (Fig. 4). In MX, fewer mutant molecules, though not significant, and significantly fewer wild-type molecules were isolated (**Figure 4A, B**). However, this did not result in a significantly different VAF (**Figure 4C**).

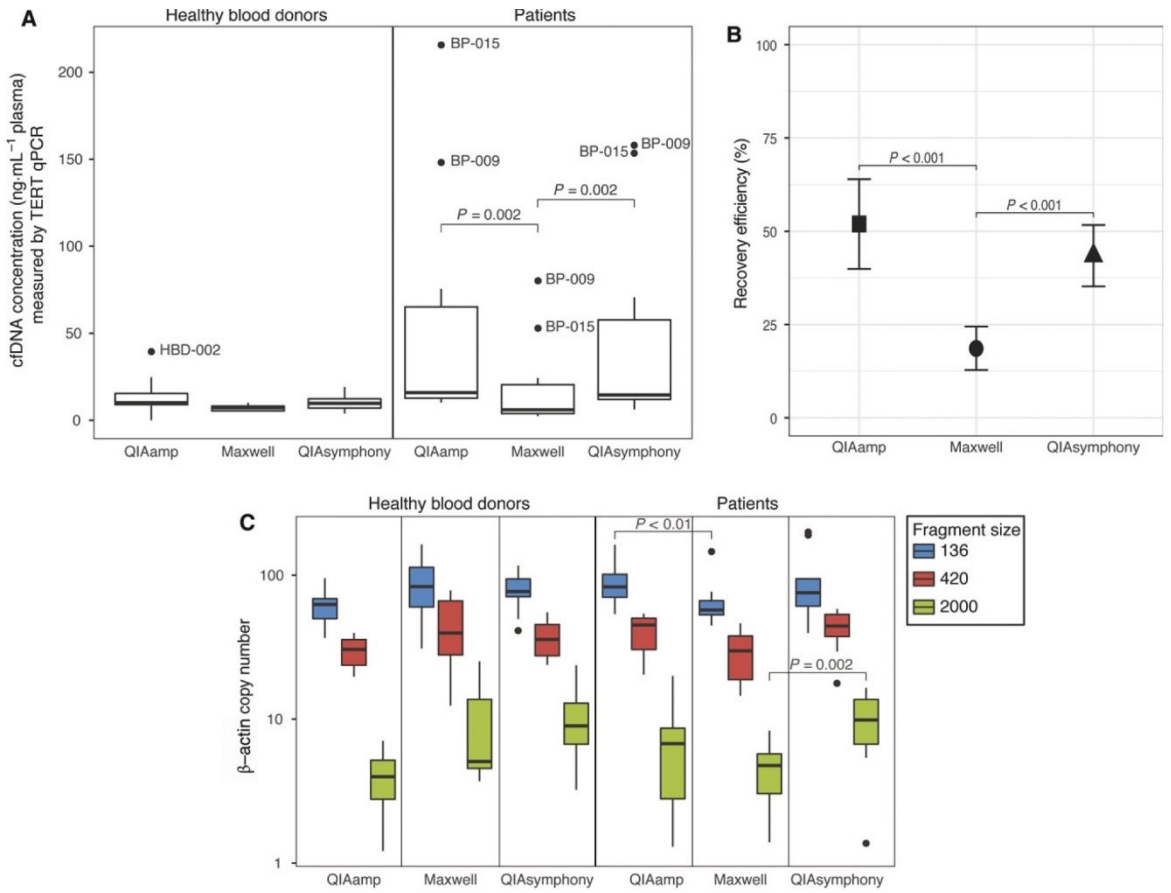


Figure 3. Effect of the different isolation platforms (QIAamp, Maxwell and QIAsymphony) on downstream cfDNA analysis.

Cell-free DNA was isolated from 2 mL matched plasma samples of healthy blood donors (N=10) and patients with metastatic cancer (N=10) and analyzed by TERT qPCR assay for cfDNA concentration (ng/mL plasma) (A), plant DNA qPCR assay (B) to determine recovery efficiency and dPCR β-actin fragmentation assay (C) to evaluate cfDNA fragment sizes. Boxes (interquartile ranges; IQR) and whiskers (1.5x IQR) are shown together with the median (black horizontal line). Outliers are indicated as single black points. Symbols ■, ● and ▲are mean values shown with whiskers (standard deviation). The Friedman test was used to test the group difference between matched samples processed by the three platforms. Significant different were post-hoc analyzed using the Wilcoxon signed rank test.

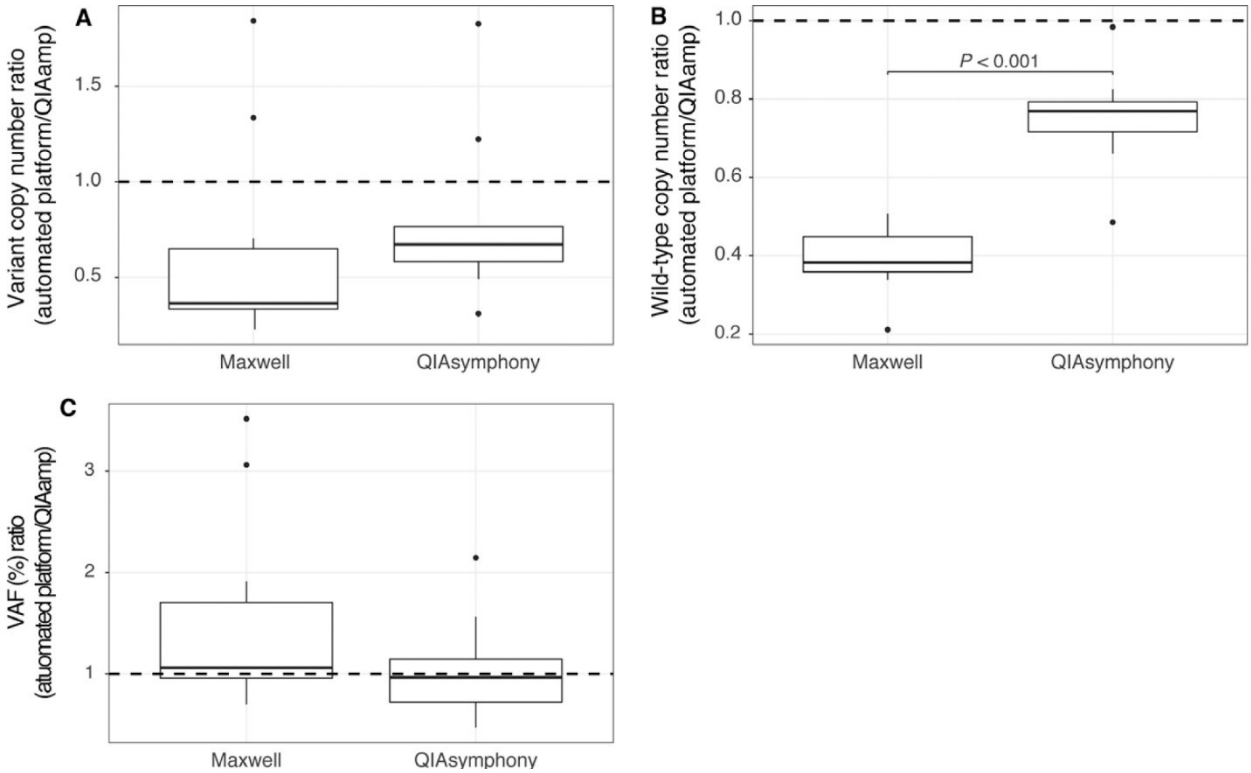


Figure 4. Somatic variant detection in patients with metastatic cancer on samples isolated with the three different isolation platforms (QIAamp, Maxwell and QIAsymphony).

Somatic variant status had been assessed in patients' primary and/or metastatic lesion as part of the standard of care. In all patients (N=10) the known somatic variant was detected in plasma isolated from the three platforms. The ratios of the mutant copy number (A), wild type copy number (B) and variant allele frequency (VAF; C) measured in the Maxwell and QIAsymphony vs. QIAamp are shown. The dashed line (ratio of 1) resembles the situation when platforms have similar results. The Wilcoxon signed rank test was used to test the difference between the platforms.

Discussion

Up to now, several studies have investigated the effect of manual and automated cfDNA isolation platforms on ctDNA quantity and quality [22-24]. However, differences in pre-analytical conditions, including plasma processing time, type of blood collection tube used, and storage conditions, hamper direct comparisons and straightforward conclusions. Here, we presented a study in which we have systematically optimized and compared automated isolation of cfDNA using QS and MX with the 'gold standard' QA.

The addition of carrier molecules like cRNA to plasma preceding cfDNA isolation increases theamount of cfDNA recovered during isolation by precipitating and binding of small molecules [25, 26]. The manual QA platform requires addition of cRNA for the standard protocol, whereas the manufacturer's protocol of both the QS and MX does not require this. In a small pilot study, we observed that the addition of cRNA to the QS protocol improved cfDNA yield, so cRNA was implemented into our standard QS protocol. However, Invitrogen has reported that cRNA might interfere with Qubit-based DNA quantification. Indeed, our findings suggest that the increase in cfDNA concentration as measured by Qubit for QS and MX is, at least in part, affected by the presence of cRNA. Data obtained from the TERT and plant DNA qPCR did not reveal any added value of cRNA to either of the automated platforms. Moreover, our fragmentation assay suggests that increasing amounts of cRNA reduce the amount of small fragments. Together, our results demonstrate that addition of cRNA to plasma does not improve cfDNA yields using these automated bead-based platforms. In our previous study using the manual QA platform, we demonstrated the superiority of CellSave tubes over EDTA tubes for collecting plasma for cfDNA/ctDNA analysis as it ensures optimal ctDNA quality when processed within 96 h after blood draw compared to only 24 h for EDTA tubes, enabling its use in multicenter clinical studies [19]. Therefore, we investigated the compatibility of CellSave tubes with QS and MX. On both platforms, we observed an increase in the isolation of large cfDNA fragments (2000 bp) in EDTA samples. This relates to the release of intact DNA from lysed leukocytes and a subsequent increase in cfDNA concentration, which we also observed here. As the recovery efficiency was not affected in CellSave tubes and the plasma samples were not contaminated with additional DNA from leukocytes, we recommend the use of Cell-Save tubes in combination with the QS or MX platform.

Currently, QA is widely used for cfDNA/ctDNA isolations, but its manual laborious and time-consuming protocol renders this method unsuitable for high throughput isolations. The competing automated platforms QS and MX both use magnetic-bead-based protocols and have comparable hands-on times. However, costs and number of samples that can be processed per run differ (**Table 2**). In HBDs, cfDNA quantity and quality were similar on all platforms. However, in patients we saw for all assays that QA and QS yielded more cfDNA than MX. As this might suggest that higher amounts of cfDNA are less efficiently isolated by the MX platform, we spiked high amounts of fragmented DNA in HBD plasma and isolated this with MX (**Figure S3**). However, these high DNA amounts were isolated efficiently by MX. Another potential explanation for the difference in performance might be the absence of proteinase K incubation step in the MX protocol. Proteinase K is used in both the QA and QS protocols and can improve cfDNA yield by inhibiting nucleases and the release of protein-bound cfDNA.

Moreover, recovery efficiency of plant DNA was lowest in MX. Altogether, this explains the lower yield of mutant and wild-type molecules isolated by MX, which may be a concern in samples with low frequent somatic variants. However, importantly, this lower yield did not translate into a significant difference in detected VAF (**Figures 4C** and **S4**). These data underline the importance of taking the used isolation method and readout (mutant molecules X mL -1 plasma or VAF) into consideration when comparing results between studies as well as for the diagnostic use of ctDNA. QS and QA performed comparable in detection of absolute numbers of mutant and wild-type molecules. Of note, other publications have observed similar performances of QA and MX in a head-to-head comparison [23, 24]. This could be related to differences in pre-analytical conditions (e.g., type of blood collection tube, plasma volume used as input), as multiple publications have demonstrated its relation to cfDNA quantity and quality [19, 27, 28]. In addition, we have optimized our QA protocol by re-eluting three times and thereby improving our cfDNA quantity. For automated magnetic- bead-based systems, this is not possible.

Conclusion

The results of this study show that the QS automated platform has comparable performance to the ‘gold standard’ QA and outperformed the MX platform depending on the readout used. The QS platform is congruent with all our predefined goals as it (a) reduces hands-on time from 180–240 to 30 min per run; (b) is able to process larger numbers of samples (96 instead of 24 at a

time); (c) isolates comparable cfDNA yield with similar efficiency; and (d) has comparable ctDNA quantity and quality to QA. Therefore, the QS can replace the more laborious QA platform, especially when high-throughput cfDNA isolation is needed.

References

1. P, M.P.a.M., *Les acides nucleiques du plasma sanguin chez l'homme.* R Seances Soc Biol Fil, 1948. **142**: p. 241–243.
2. Stroun, M., et al., *Neoplastic characteristics of the DNA found in the plasma of cancer patients.* Oncology, 1989. **46**(5): p. 318-22.
3. Bidard, F.C., et al., *Detection rate and prognostic value of circulating tumor cells and circulating tumor DNA in metastatic uveal melanoma.* Int J Cancer, 2014. **134**(5): p. 1207-13.
4. Dawson, S.J., et al., *Analysis of circulating tumor DNA to monitor metastatic breast cancer.* N Engl J Med, 2013. **368**(13): p. 1199-209.
5. Diaz, L.A., Jr. and A. Bardelli, *Liquid biopsies: genotyping circulating tumor DNA.* J Clin Oncol, 2014. **32**(6): p. 579-86.
6. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics.* Nat Med, 2008. **14**(9): p. 985-90.
7. Forshew, T., et al., *Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA.* Sci Transl Med, 2012. **4**(136): p. 136ra68.
8. Herbreteau, G., et al., *Quantitative monitoring of circulating tumor DNA predicts response of cutaneous metastatic melanoma to anti-PD1 immunotherapy.* Oncotarget, 2018. **9**(38): p. 25265-25276.
9. Murtaza, M., et al., *Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA.* Nature, 2013. **497**(7447): p. 108-12.
10. Pugh, T.J., *Circulating Tumour DNA for Detecting Minimal Residual Disease in Multiple Myeloma.* Semin Hematol, 2018. **55**(1): p. 38-40.
11. Shinozaki, M., et al., *Utility of circulating B-RAF DNA mutation in serum for monitoring melanoma patients receiving biochemotherapy.* Clin Cancer Res, 2007. **13**(7): p. 2068-74.
12. Oxnard, G.R., et al., *Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA.* Clin Cancer Res, 2014. **20**(6): p. 1698-1705.
13. Sefrioui, D., et al., *Clinical value of chip-based digital-PCR platform for the detection of circulating DNA in metastatic colorectal cancer.* Dig Liver Dis, 2015. **47**(10): p. 884-90.
14. Zill, O.A., et al., *Cell-Free DNA Next-Generation Sequencing in Pancreatobiliary Carcinomas.* Cancer Discov, 2015. **5**(10): p. 1040-8.

15. Fleischhacker, M. and B. Schmidt, *Circulating nucleic acids (CNAs) and cancer--a survey.* Biochim Biophys Acta, 2007. **1775**(1): p. 181-232.
16. Bettegowda, C., et al., *Detection of circulating tumor DNA in early- and late-stage human malignancies.* Sci Transl Med, 2014. **6**(224): p. 224ra24.
17. Elshimali, Y.I., et al., *The clinical utilization of circulating cell free DNA (CCFDNA) in blood of cancer patients.* Int J Mol Sci, 2013. **14**(9): p. 18925-58.
18. Jahr, S., et al., *DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells.* Cancer Res, 2001. **61**(4): p. 1659-65.
19. van Dessel, L.F., et al., *Application of circulating tumor DNA in prospective clinical oncology trials - standardization of preanalytical conditions.* Mol Oncol, 2017. **11**(3): p. 295-304.
20. Kang, Q., et al., *Comparative analysis of circulating tumor DNA stability In K3EDTA, Streck, and CellSave blood collection tubes.* Clin Biochem, 2016. **49**(18): p. 1354-1360.
21. Invitrogen, *Qubit dsDNA assay specificity in the presence of single-stranded DNA. Application note.* 2016.
22. Devonshire, A.S., et al., *Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification.* Anal Bioanal Chem, 2014. **406**(26): p. 6499-512.
23. Perez-Barrios, C., et al., *Comparison of methods for circulating cell-free DNA isolation using blood from cancer patients: impact on biomarker testing.* Transl Lung Cancer Res, 2016. **5**(6): p. 665-672.
24. Sorber, L., et al., *A Comparison of Cell-Free DNA Isolation Kits: Isolation and Quantification of Cell-Free DNA in Plasma.* J Mol Diagn, 2017. **19**(1): p. 162-168.
25. Kishore, R., et al., *Optimization of DNA extraction from low-yield and degraded samples using the BioRobot EZ1 and BioRobot M48.* J Forensic Sci, 2006. **51**(5): p. 1055-61.
26. Shaw KJ, T.L., Docker PT, Dyer CE, Greenman J, Greenway GM and Haswell SJ, *The use of carrier RNA to enhance DNA extraction from microfluidic-based silica monoliths.* Anal Chim Acta, 2009. **652**: p. 231–233.
27. Haselmann, V., et al., *Results of the first external quality assessment scheme (EQA) for isolation and analysis of circulating tumour DNA (ctDNA).* Clin Chem Lab Med, 2018. **56**(2): p. 220-228.

28. Volckmar AL, S.H., Riediger A, Fioretos T, Schirmacher P, Endris V, Stenzinger A and Dietz S, *field guide for cancer diagnostics using cellfree DNA: From principles to practice and clinical applications.* Genes Chromosom Cancer, 2018. **57**: p. 123–139.

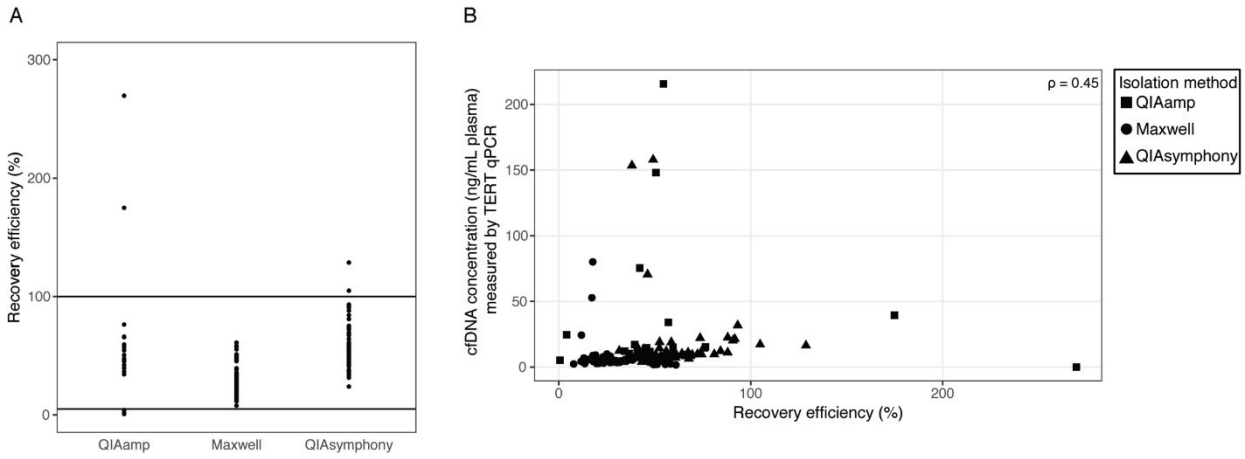


Figure S1. Overview of the recovery efficiency of synthetic plant DNA in all samples isolated with the different platforms (QIAamp, Maxwell and QIAsymphony).

(A) Dot plot of the recovery efficiency for each isolation platform, as analyzed by dPCR using spiked-in synthetic plant DNA. Samples it recovery efficiency < 5% or > 100% (black horizontal line) were excluded from the analysis. (B) Correlation between recovery efficiency and cfDNA concentration (ng/mL plasma) measured by TERT qPCR assay. Correlations were tested by Spearman's rank correlation coefficient. *P<0.001.

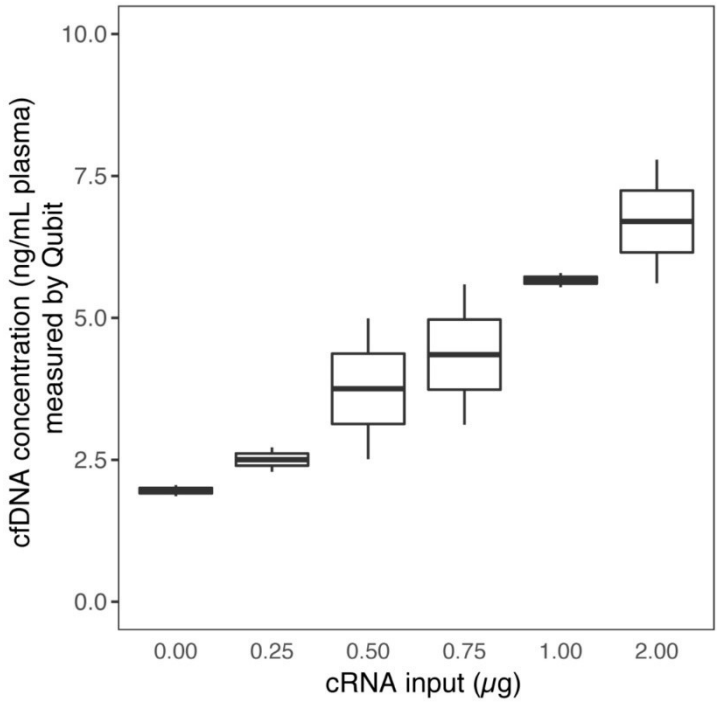


Figure S2. Effect of cRNA addition on cfDNA quantity using the QIAsymphony platform.

cfDNA concentration (ng/mL plasma) was determined by Qubit after adding increasing amounts of cRNA (0 - 4 µg) before start of the plasma isolation. Boxes (interquartile ranges; IQR) and whiskers (1.5x IQR) are shown together with the median (black horizontal line).

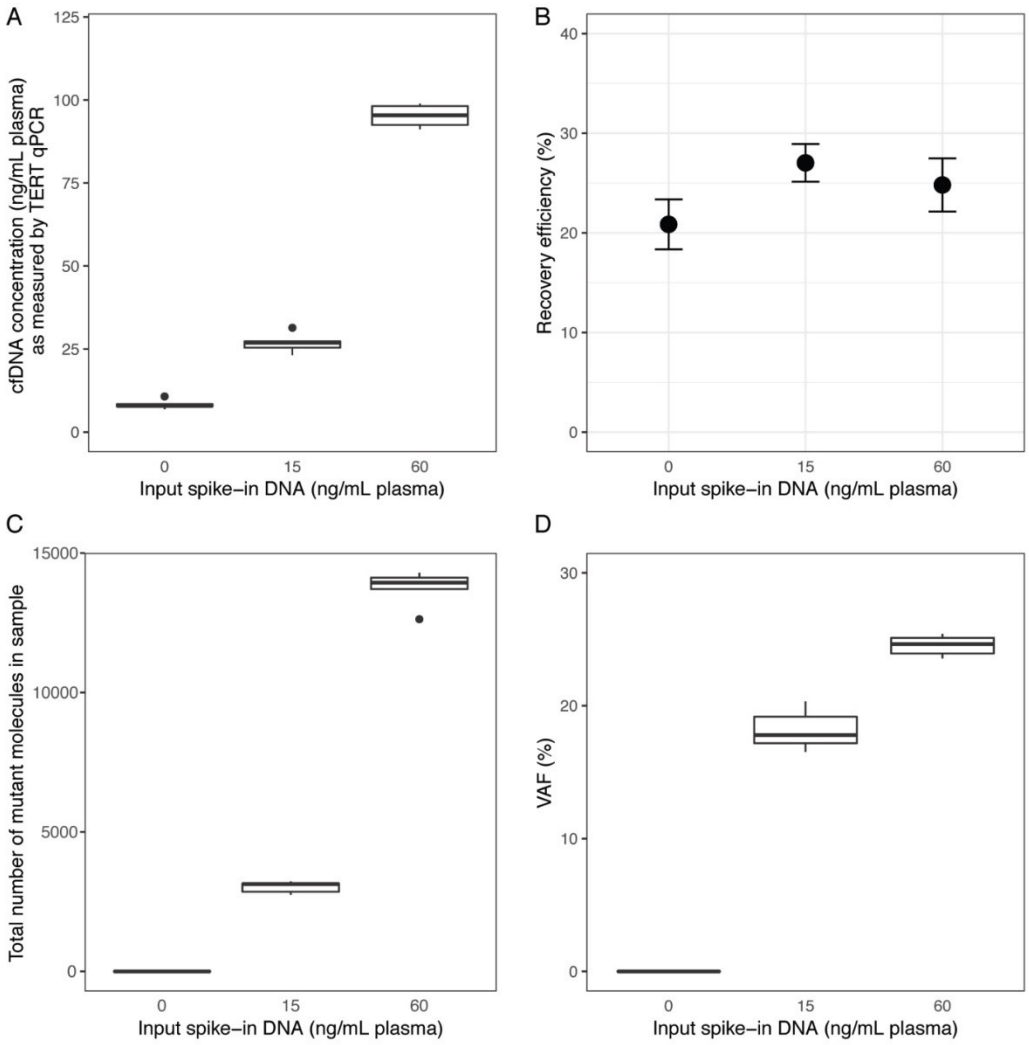


Figure S3. Performance of the Maxwell platform using increasing DNA input (0, 15 and 60 ng/mL fragmented cell line DNA has been spiked in healthy blood donor plasma).

The effect on cfDNA concentration (ng/mL plasma) measured by TERT qPCR (A), recovery efficiency measured by plant DNA qPCR (B) and total number of mutant molecules (C) and variant allele frequency (VAF) (D) are shown. Boxes (interquartile ranges; IQR) and whiskers (1.5x IQR) are shown together with the median (black horizontal line). Outliers are indicated as single black points. Symbol ● is mean value shown with whiskers (standard deviation). *N=5*

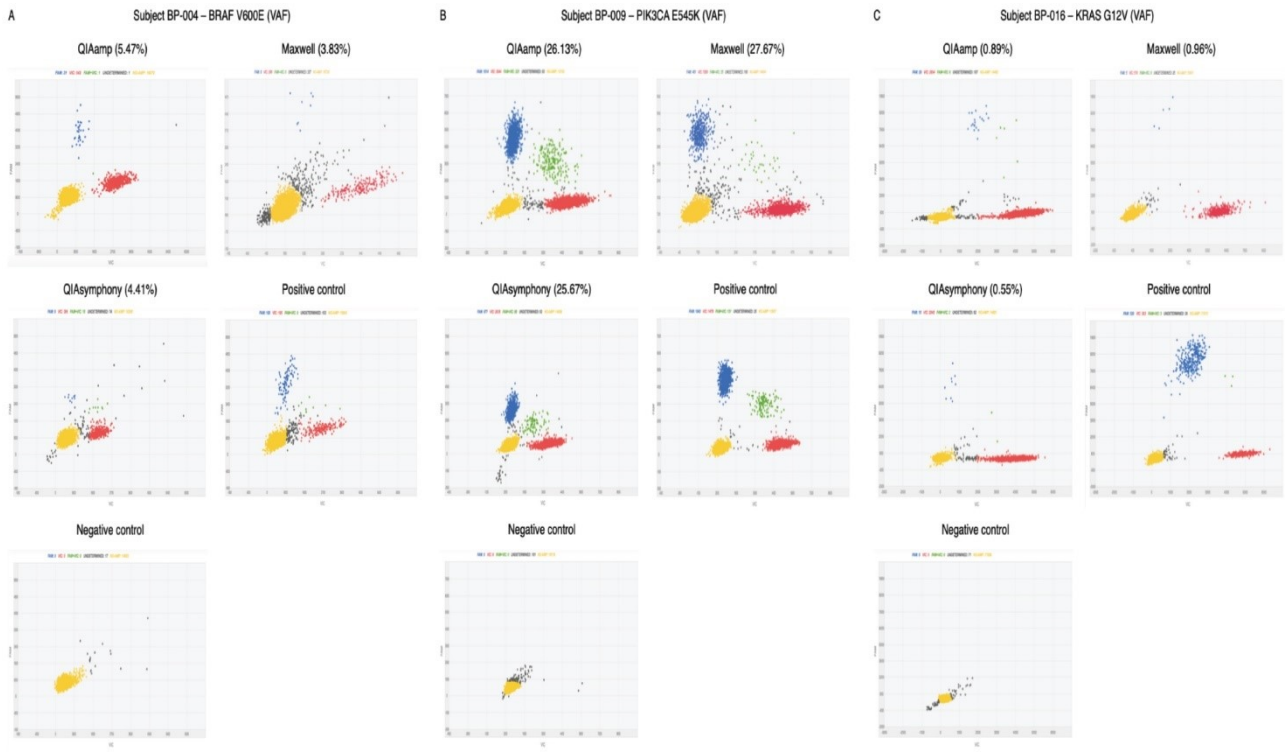


Figure S4. Representative data images of SNP genotyping dPCR assay isolated with the different platforms (QA, MX and QS).

A subject with an intermediate (A), high (B) and low (C) variant allele frequency (VAF) are shown. On the Y-axis, positive FAM signal represents mutant molecules (blue dots); on the X-axis, positive VIC signal represents wild type molecules (red dots). Green dots reflect the presence of a mutant and a wild type molecule in a single well.

Table S1. Custom primer and probe sequences used for qPCR.

Assay Name	Forward primer	Reverse primer	Probe	Amplicon size (bp)	Manufacturer
β-actin 136 bp	5'-GCG CCG TTC CGA AAG TT-3'	5'- CGG CGG ATC GGC AAA -3'	FAM-ACC GCC GAG ACC GCG TC-MGBNFQ	136	Invitrogen
β-actin 420 bp	5'-CCG CTA CCT CTT CTG GTG-3'	5'-GAT GCA CCA TGT CAC ACT G-3'	VIC-CCT CCC TCC TTC CTG GCC TC-BHQ	420	Invitrogen
β-actin 2000 bp	*The β-actin 2000 bp fragment is detected when both primers and probes of the β-actin 136 bp assay and the β-actin 420 bp are able to amplify a long DNA fragment. This double positive signal is detected by the QuantStudio 3D Digital PCR System (van Dessel et al., 2017).*				Invitrogen
Plant	5'-GAT CTT CAA CCA GGA GAT CA-3'	5'-AGT GAC AGT GAG GAC AAT CC-3'	FAM-ACC CAT CTT CAC CGG A-BHQ1	70	Primers: IDT Probe: Sigma-Aldrich, Saint Louis, Missouri, USA

Table S2. Standard SNP genotyping assays.

Assay ID	Assay Name	Gene	Cosmic ID	Amino acid change	Nucleotide change
AHS1P6Q	NRAS_584	*NRAS*	584	p.Q61R	c.182A>G
AHD2BW0	KRAS_532	*KRAS*	532	p.G13D	c.38G>A
AHABHHX	PIK3CA_763	*PIK3CA*	763	p.E545K	c.1633G>A

Table S3. Custom SNP genotyping assays.

Assay ID	Assay Name	Gene	Cosmic ID	Amino acid change	Nucleotide change	Forward primer	Reverse primer	Probe	Amplicon size (bp)	Manufacturer
ANNKR4W	BRAF_ V600K_72bp	*BRAF*	473	V600K	c.1798_1799GT>AA	5′-TCA TGA AGA CCT CAC AGT AAA AAT AGG T-3′	5′- TGG GAC CCA CTC CAT CGA-3′	*Variant*: FAM-TGG TCT AGC TAC AAA GA-NFQ *Wild type*: VIC- TTT TGG TCT AGC TAC AGT GA-NFQ	72	ThermoFisher Scientific
AN47WF2	BRAF_ V600E_72bp-2	*BRAF*	475	V600E	c.1799_1800 TG>AA	5′-TCA TGA AGA CCT CAC AGT AAA AAT AGG T-3′	5′-TGG GAC CCA CTC CAT CGA-3′	*Variant*: FAM-TTG GTC TAG CTA CAG AAA-NFQ *Wild type*: VIC- TTT TGG TCT AGC TAC AGT GA-NFQ	72	
AN9HJKW	KRAS_ G12C_76bp	*KRAS*	516	G12C	c.34G>T	5′-TGC TGA AAA TGA CTG AAT ATA AAC TTG TG-3′	5′- AGC TGT ATC GTC AAG GCA CTC TT-3′	*Variant*: FAM-TTG GAG CTT GTG GCG TA-NFQ *Wild type*: VIC- TTG GAG CTG GTG GCG T-NFQ	76	
ANU63FK	KRAS_ G12D_76bp	*KRAS*	521	G12D	c.35G>A	5′-TGC TGA AAA TGA CTG AAT ATA AAC TTG TG-3′	5′-AGC TGT ATC GTC AAG GCA CTC TT-3′	*Variant*: FAM-TTG GAG CTG TTG GCG TA-NFQ *Wild type*: VIC-TTG GAG CTG GTG GCG T-NFQ	76	
ANAAAYM	KRAS_ G12V_76bp	*KRAS*	520	G12V	c.35G>T	5′-TGC TGA AAA TGA CTG AAT ATA AAC TTG TG-3′	5′-AGC TGT ATC GTC AAG GCA CTC TT-3′	*Variant*: FAM-TTG GAG CTG ATG GCG TA-NFQ *Wild type*: VIC-TTG GAG CTG GTG GCG T-NFQ	76	

Chapter 4

TP53 mutations in serum circulating cell-free tumor DNA as longitudinal biomarker for high-grade serous ovarian cancer

Vitale SR, Groenendijk FH, van Marion R, Beaufort C, Helmijr J, Jan Dubbink H, Dinjens WNM, Ewing-Graham PC, Smolders R, van Doorn LHC, Boere IA, Berns EMJJ, Helleman J, Jansen MPHM

Biomolecules 2020 Mar 7;10(3):415
doi: 10.3390/biom10030415

Abstract

The aim of this study was to determine an optimal workflow to detect TP53 mutations in baseline and longitudinal serum cell free DNA (cfDNA) from high-grade serous ovarian carcinomas (HGSOC) patients and to define whether *TP53* mutations are suitable as biomarker for disease.

TP53 was investigated in tissue and archived serum from 20 HGSOC patients by a next-generation sequencing (NGS) workflow alone or combined with digital PCR (dPCR). AmpliSeq™-focused NGS panels and customized dPCR assays were used for tissue DNA and longitudinal cfDNAs, and Oncomine NGS panel with molecular barcoding was used for baseline cfDNAs.

TP53 missense mutations were observed in 17 tissue specimens and in baseline cfDNA for 4/8 patients by AmpliSeq, 6/9 patients by Oncomine, and 4/6 patients by dPCR. Mutations in cfDNA were detected in 4/6 patients with residual disease and 3/4 patients with disease progression within six months, compared to 5/11 patients with no residual disease and 6/13 patients with progression after six months. Finally, mutations were detected at progression in 5/6 patients, but not during chemotherapy.

NGS with molecular barcoding and dPCR were most optimal workflows to detect *TP53* mutations in baseline and longitudinal serum cfDNA, respectively. *TP53* mutations were undetectable in cfDNA during treatment but re-appeared at disease progression, illustrating its promise as a biomarker for disease monitoring.

Introduction

Epithelial ovarian cancer is the most lethal malignancy among gynecological cancers in the Western world, partly due to the advanced disease stage at the time of diagnosis in most patients [1]. A large proportion of these patients will have a recurrence or progression within two years and ultimately die of their disease. Only 30% of women initially diagnosed with advanced-stage disease will survive more than five years. However, when the disease is diagnosed while still confined to the ovary, five-year survival is 70% to 90% [2].

Circulating tumor DNA (ctDNA) and cell-free DNA (cfDNA) isolated from blood have recently been extensively investigated as potential blood-based biomarkers for several cancer types [3–5]. In ovarian cancer patients, higher levels of cfDNA were found as compared to healthy donors [6,7] and patients with benign disease [1]. Higher levels of cfDNA were associated with advanced disease stage, high grade, and poorer prognosis [1,8]. In addition, it has been reported in an orthotopic mouse model that progression of disease could be monitored by measuring human cfDNA. In this model, the amount of cfDNA correlated significantly with tumor weight [9]. However, increased level of cfDNA can also be found in patients with benign lesions, inflammatory disease, and tissue trauma and is therefore not a specific tumor biomarker [10]. The presence of tumor-specific genetic alterations in cfDNA could potentially offer a more specific approach.

In this context, it is of interest that 96% of high-grade serous ovarian carcinomas (HGSOC), representing the majority of advanced stage ovarian cancers, have a mutated *TP53* gene [11,12]. Therefore, the presence of *TP53* mutations in cfDNA could potentially be used as a tumor-specific marker for HGSOC. However, *TP53* mutations can also be detected in plasma DNA from healthy, especially older, individuals and patients with other tumor types [13]. We hypothesized that tumor-specific *TP53* missense mutations in archived serum cfDNA of HGSOC patients could be used as a biomarker at baseline, and as a marker to monitor tumor load during therapy and thereafter. This hypothesis is supported by data from a retrospective study of 40 HGSOC patients, which showed that plasma circulating tumor DNA correlated with disease load at start of therapy. Furthermore, a decrease of less than 60% in *TP53* mutation frequency after one cycle of chemotherapy was associated with shorter time to progression (TTP) [14]. However, to date, it is not clear which technique is optimal for detecting and monitoring *TP53* mutations in cfDNA.

In the current study, we evaluate different NGS workflows, used alone and combined with digital PCR (dPCR), for detecting *TP53* mutations in minute amounts of archived serum (<1 mL). These samples were taken at the time of diagnosis from 20 patients with advanced HGSOC patients. The amount of baseline cfDNA with *TP53* mutation was correlated with residual disease after debulking surgery and progression-free survival after platin-based chemotherapy. We also analyzed mutations in serum taken during chemotherapy and at disease progression in the same group of patients to investigate the potential of *TP53* mutations in blood as a marker of disease progression.

Materials and Methods

Study Design and Patient Characteristics

For this retrospective study, we selected 20 patients with FIGO stage IIIC–IV ovarian cancer diagnosed at the Erasmus MC Cancer Institute in Rotterdam, The Netherlands. This retrospective study of archived tissue and serum was approved by the medical ethics committee of the Erasmus MC Rotterdam, the Netherlands (MEC-2002-949; MEC-2008-183) and performed in accordance with the principles of the Declaration of Helsinki and the local law. The study was carried out according the REMARK guidelines and Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (https://www.federa.org/codes-conduct). The majority of patients had high-grade serous ovarian cancer (N = 19), only one with adenocarcinoma (N = 1). Most patients received primary or interval debulking surgery, and all received platinum-based chemotherapy. Patient subsets were defined based on size of residual disease (optimal debulking defined as 0–1 cm vs. non-optimal debulking defined as >1 cm residual disease) and progression-free survival (PFS ≤6 months vs. PFS >6 months). The study evaluated two workflows (**Table 1**): Workflow I evaluated only *TP53* by NGS and dPCR on tissue DNA and longitudinal serum cfDNAs, whereas workflow II applied only NGS using multigene panels including *TP53* on tissue DNA and baseline cfDNA. Patient, clinical characteristics, and workflow details are summarized in **Tables 1** and **2**, respectively.

Table 1. Study design and clinical characteristics.

	Study Design Details			Clinical Characteristics			
Patient	**Workflow**	**Sample Subset**	**Age**	**FIGO Stage**	**Debulking Surgery ***	**Residual Disease**	**Progression-Free Survival (PFS)**
1	**NGS & dPCR:** *TP53* only	Tissue & longitudinal cfDNA	43	IV	PDS	0–1 cm	>6 months
2			37	IIIC	PDS	1 cm or more	>6 months
3			69	IIIC	PDS	1 cm or more	>6 months
4			56	IV	PDS	0–1 cm	>6 months
5			73	IIIC	IDS	0–1 cm	>6 months
6 **			53	IIIC	PDS	0–1 cm	>6 months
7			58	IV	PDS	1 cm or more	>6 months
8			58	IIIC	PDS	0–1 cm	>6 months
9			72	IIIC	IDS	1 cm or more	>6 months
10 **			75	IV	IDS	1 cm or more	>6 months
11	**NGS only:** Multigene panels including *TP53*	Tissue & baseline cfDNA	47	IV	IDS	0–1 cm	>6 months
12			47	IIIC	IDS	0–1 cm	>6 months
13			57	IIIC	IDS	0–1 cm	>6 months
14			49	IC	PDS	0–1 cm	>6 months
15			61	IV	IDS	0–1 cm	>6 months
16			63	IIIC	IDS	0–1 cm	0–6 months
17			45	IV	IDS	0–1 cm	0–6 months
18 **			65	IV	NA *	1 cm or more	0–6 months
19			47	IV	NA *	1 cm or more	0–6 months
20			72	IV	IDS	1 cm or more	0–6 months

* PDS = primary Debulking surgery, IDS = interval Debulking surgery, NA = not available; ** Patients were excluded from cfDNA analysis because their tissue had no TP53 missense mutations.

Table 2. Cell-free DNA (cfDNA) workflow details.

	Worflow I: NGS & dPCR—*TP53* only		Worflow II: NGS only—Multigene Hotspot Panels Including *TP53*	
Samples	Tissue DNA	Longitudinal cfDNA	Tissue DNA	Baseline cfDNA only
Panel details	24 amplicons, 2550 bp		328 amplicons, 35793 bp, 41 genes	26 amplicons, 4420 bp, 10 genes
Input amount	10 ng	1.5–3.3 ng	10 ng	15–20 ng
Mean reads depth coverage (range)	1725 reads/amplicon (943–3584 reads)	1851 reads/amplicon (843–3776 reads)	1166 reads/amplicon (430–1796 reads)	32902 reads/amplicon (5786–59976 reads)
Estimated costs per sample for:				
NGS	€150–€250	€150–€250	€250–€350	€350–€450
dPCR (€400 per mutation assay)	€20–€30	€20–€30	NA	NA
Workflow cfDNA NGS vs. dPCR total costs for:				
Baseline cfDNA only		€150–€250 vs. €420–€430		€350–€450 vs. €420–€430
cfDNA at baseline & progression		€300–€500 vs. €440–€460		€700–€900 vs. €440–€460
3 longitudinal cfDNAs		€450–€750 vs. €460–€490		€1050–€1350 vs. €460–€490
5 longitudinal cfDNAs		€750–€1250 vs. €500–€550		€1750–€2250 vs. €500–€550

DNA Isolation

The DNA was extracted from fresh frozen tumor tissue specimens (N = 8) and from formalin-fixed paraffin embedded (FFPE) tissue (N = 12) taken at debulking surgery as described previously [15]. Tissues were sectioned for DNA isolation and the percentage of tumor cells was evaluated in a hematoxylin-eosin-stained section as described earlier [16]. The cfDNA was isolated from archived minute amounts of serum samples taken at diagnosis for all patients in both workflows, and at chemotherapy and at disease progression for patients of workflow I. The QIAamp Circulating Nucleic Acid kit (Qiagen, KJ Venlo, The Netherlands) was used to isolate cfDNA from a median of 400 µL serum (range 100–1000 µL) according to the manufacturer's manual. This

cfDNA was isolated into 20 µL elution buffer. A Qubit® 2.0 fluorimeter (Thermo Scientific, Carlsbad, California, USA) and the Quant-iT dsDNA high-sensitivity assay (Invitrogen, Life Technologies, Carlsbad, CA, USA) were used to quantify the isolated DNA yields and concentrations.

Next-Generation Sequencing

Different NGS panels and sequencer platforms were used in the two workflows all purchased from Life Technologies (Carlsbad, CA, USA). The panels differ in amplicon and target size, methodology, and sequencing costs per sample (**Table 2**). Tissue DNA was sequenced on the Ion Torrent Personal Genome Machine (ion-PGM) using two Ion Ampliseq focused panels. In workflow I, the Ion AmpliSeq™ *TP53* community panel (24 amplicons, 2550 bp; analyzing exons and UTRs of the TP53 gene) was used. In workflow II, a customized Ion Ampliseq Diagnostic V5.1 targeted panel (328 amplicons, 35,793 bp, 41 genes) was applied to sequence all *TP53* exons. Serum cfDNA was sequenced with the Ion AmpliSeq™ *TP53* community panel on the ion-PGM (workflow I) or with the Oncomine breast cfDNA NGS assay with molecular barcoding (26 amplicons, 4420 bp, 10 genes) on an Ion S5XL sequencer (workflow II). The ampliseq panels have a limit of detection (LOD) for mutation frequencies of at least 1%, whereas the Oncomine panel has a LOD of 0.1% at 20 ng DNA input. For this lower LOD, much deeper read depth coverage is needed for the Oncomine panel (>20k coverage) than Ampliseq panels (<1k coverage). Consumables, kits, software packages, and protocols for the NGS analyses with Ampliseq and Oncomine focused panels were used as indicated by the manufacturer and as previously described by us [3,16,17]. Briefly, 10 ng tumor DNA was used for all patients as input in the Ampliseq library preparation. For cfDNA library preparations, at least 1.5 ng cfDNA was used with equal amounts for all three sera per patient for the Ampliseq panel in workflow I (median 2.2 ng; range per patient: 1.5–3.3 ng) and at least 15 ng up to 20 ng of cfDNA was used for the Oncomine panel in workflow II. Samples were sequenced on Ion 318 and 530 chips for workflows I and II, respectively. Sequencing with Ampliseq panels was performed with on average tumor tissue DNA reads depth coverage of 1725 reads/amplicon (range: 943–3584 reads) for workflow I and 1166 reads/amplicon (range: 430–1796 reads) for workflow II. The average cfDNA reads depth coverage was 1851 reads/amplicon (range: 843–3776 reads) for workflow I and 32,902 reads/amplicon (range: 5786–59,976 reads) for workflow II.

Bio-Informatics for SNV Detection and Evaluation

The Torrent Suite v4.0 (Thermo Scientific, Carlsbad, California, USA) was used for raw data analyses, base calling, and alignment. Variant Caller v4.4.2.1 (VC, Thermo Scientific, Carlsbad, California, USA) was applied to detect DNA sequence alterations. Annotation of the variants was performed by a custom pipeline including ANNOVAR (openbioinformatics.org/annovar) in a Galaxy (galaxyproject.org) environment. For the initial VC analysis of each tumor and serum DNA-sample, somatic low-stringency filter settings were applied to detect DNA variants when compared to the reference genome (hg19; build 37). For each sample, only sequences with 100 reads depth or more were evaluated. Subsequently, *TP53* mutations were visually examined using Integrative Genomics Viewer (IGV, California, USA) software (http//www.broadinstitute.org/igv).

Digital PCR

Independent validation of six identified *TP53* mutations were performed using TaqMan® SNP genotyping assays on the QuantStudio™ 3D Digital PCR system (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer specifications. All the assays, except *TP53* p.R282W, were designed in-house using the Thermo Fisher Custom TaqMan® Assay Design Tool and ordered as Custom TaqMan® SNP Genotyping assay from Thermo Fisher (**Tables S1** and **S2**). The PCR reaction mix was prepared in a final volume of 17.4 µL containing 30 ng of tumor DNA or ranging from 3.6 ng to 20 ng for serum cfDNA. Then, the amplification mix was partitioned into ~20,000 wells loaded in a QuantStudio 3D Digital PCR Chip v2 and run on a ProFlex 2x Flat PCR System. The temperature profile for amplification was: an activation step of 10 min at 96 °C, followed by 40 cycles of 2 min at 60 °C, 30 s incubation at 98 °C, 2 min at 60 °C, and pause at 10 °C. The QuantStudio™ 3D analysisSuite™ was used to analyze the end-point fluorescence data to determine the proportion of templates, with and without a mutation, and to calculate the Mutation Allele Frequency (MAF). At least one negative and one positive control were added to each run.

TP53 Immunohistochemistry

FFPE tumor tissue sections of patients 1 and 2 were stained with polyclonal p53 antibodies clone DO-1 (1:200, Santa Cruz Biotechnogy, Heidelberg, Germany) and DO-7 (1:100, Dako, Santa Clara, USA) as described previously [18]. The DO-1 and DO-7 recognize overlapping epitopes. For patients 3–20, mouse monoclonal p53 antibody Bp53-11 (Ventana 769-2541, Ventana Medical Systems, Roche, Tuscon, AZ, USA) was used. Detection was performed using Ventana Benchmark Ultra detective with Ultraview Universal DAB detection kit (Ventana 760-500, Ventana Medical Systems, Roche, Tuscon, AZ, USA) and antigen retrieval Cell Conditioning Solution (CC1) (Ventana 950-124, Ventana Medical Systems, Roche, Tuscon, AZ, USA). The p53 expression was scored as previously described by Kobel et al. [19] and categorized into overexpression, complete absence, cytoplasmic, or wild-type.

Statistics

The study complied with reporting recommendations for tumor marker prognostic studies (REMARK) criteria [20]. Samples were called positive for TP53 non-synonymous mutations when the mutation frequency was above 1% when detected by Ampliseq panels or above 0.1% when detected by Oncomine or digital PCR assays. Statistical analyses were performed with Microsoft and Simple Interactive Statistical Analysis (SISA; http://www.quantitativeskills.com/sisa/index.htm). The analyses included student t-tests for continuous variables, Chi-square test for categorical variables. p-values were two-sided, and significance was defined at <0.05.

Results

***TP53* Mutation and Protein Expression Analysis in Tumor Tissue**

First the *TP53* mutation status and protein expression were analyzed in tumor tissues. As *TP53* mutations in our archived serum not only might originate from tumor cells, but can also arise from clonal hematopoiesis, we first defined tumor-specific *TP53* mutations in tissue by NGS and immunohistochemistry. Missense *TP53* mutations and strong nuclear p53 protein expression were detected in all but three patients with no or synonymous mutations (**Figure 1, Table 3**).

Only non-synonymous *TP53* mutations with (aberrant) nuclear staining in tumor tissue were then evaluated in serum cfDNA.

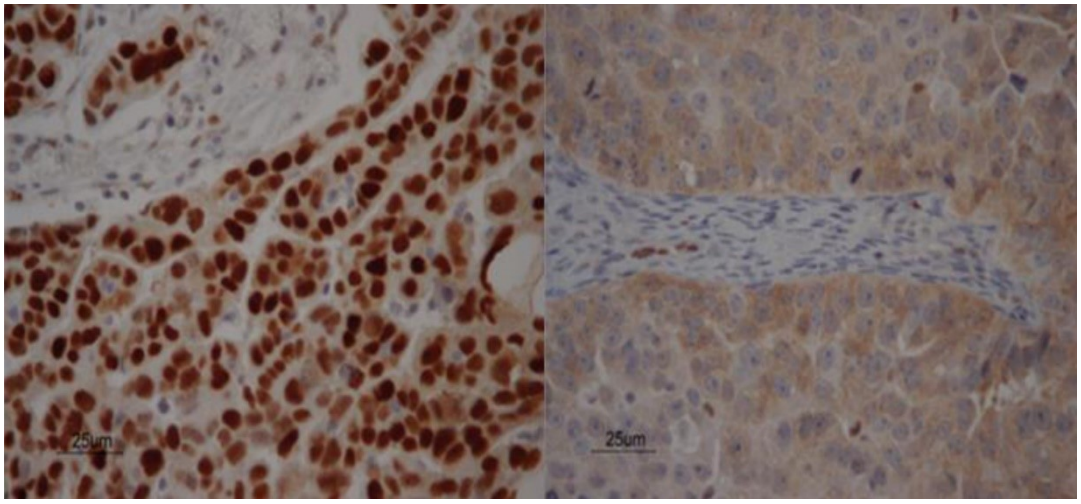


Figure 1. Tissue p53 staining.

Two examples of immunohistochemical localization of p53 expression in patients with advanced stage ovarian cancer. Staining of sporadic nuclei with p53 antibody is seen in the stroma in both figures, acting as internal control. The p53 expression showed strong nuclear staining in patient 5 with a *TP53* p.K132R mutation (left figure) and cytoplasmic staining in patient 6 without a *TP53* mutation (right figure).

Table 3. Tumor tissue and serum characteristics.

					cfDNA Yields per mL Serum (ng/mL)				cfDNA MAF (NGS)		cfDNA MAF (dPCR)			
Subset	**Patient**	**Identified *TP53* Mutation in Tissue**	**MAF Tissue**	**TP53 IHC**	**D**	**C**	**P**	**Excluded in cfDNA Analysis**	**D**	**P**	**D**	**P**	**Additional Mutation(s)**	**MAF**
	1	p.Y163C	57%	nuclear ++	17	16	56		0.0%	0.0%	0.0%	0.8%		
	2	p.C275Y	27%	nuclear ++	30	56	44		1.0%	0.0%	5.3%	0.1%		
	3	p.P151R	73%	nuclear ++	148	30	55		0.0%	6.0%	NE	NE		
	4	p.R282W	78%	nuclear ++	37	39	64		30.0%	0.0%	31.9%	0.4%		
I. NGS & dPCR:	5	p.K132R	21%	nuclear ++	35	40	35		1.0%	1.0%	2.6%	1.8%		
TP53 only	6	No mutation	-	cytoplasmatic	52	28	46	no mutation	-	-	-	-		
	7	p.Y163C	64%	nuclear ++	56	46	39		2.0%	1.0%	6.2%	2.9%		
	8	p.C275Y	7%	nuclear ++	119	35	84		0.0%	0.0%	0.0%	0.0%		
	9	p.C277F	58%	nuclear ++	111	44	275		0.0%	0.0%	NE	NE		
	10	p.N131N	3%	cytoplasmatic	229	106	338	synonymous	-		NE	NE		
	11	p.E286G	81%	nuclear ++	174				1.9%					
	12	p.Y205D	51%	nuclear ++	174				0.0%					
	13	p.F134V	75%	nuclear ++	102				0.0%				*TP53* p.T253I	0.1%
II. NGS only:	14	p.K132R	79%	nuclear ++	265				0.0%					
Multigene	15	p.L194R	88%	nuclear ++	83				0.3%					
hotspot panels	16	p.Y220C	65%	nuclear ++	98				0.0%				*ESR1* p.R394S	0.3%
including *TP53*	17	p.C176W	56%	nuclear ++	71				25.6%					
	18	Unknown (p.R280G)*[1]	88%	N/A	628				8.74%					
	19	p.E258G	85%	nuclear ++	264				0.6%					
	20	p.P278S	60%	nuclear ++	107				0.5%				*PIK3CA* p.H1047R	1.3%

IHC = immunohistochemistry, D = at diagnosis, C = after chemotherapy, P = at disease progression, NGS = next generation sequencing, dPCR = digital PCR, MAF = mutant allele frequency, *unknown = mutation was known, but no tissue was available, N/A = not available, ++ = high TP53 levels nuclear staining.

[1] the table is a slightly modified version of which is published.

Serum cfDNA Yields at Diagnosis and Over Time

The cfDNA yield per mL serum showed a range between 16 and 338 ng for workflow I and between 71 and 628 ng for workflow II (Table 3). Median cfDNA yields at diagnosis were comparable between non-optimal debulked patients (130ng/mL (N = 8)) and optimally debulked patients (90 ng/mL (N = 12); p = 0.218), and between patients with PFS shorter vs. longer than six months (median 107 (N = 5) and 102 ng/mL (N = 15); p = 0.30) (**Figure 2A, B**). The longitudinal serum cfDNA yields in workflow I were overall lower during chemotherapy (median: 39 ng/mL; p = 0.108) but similar at disease progression (median: 55 ng/mL; p = 0.627) when compared to cfDNA yields at diagnosis (median: 54 ng/mL) (**Figure 2C**).

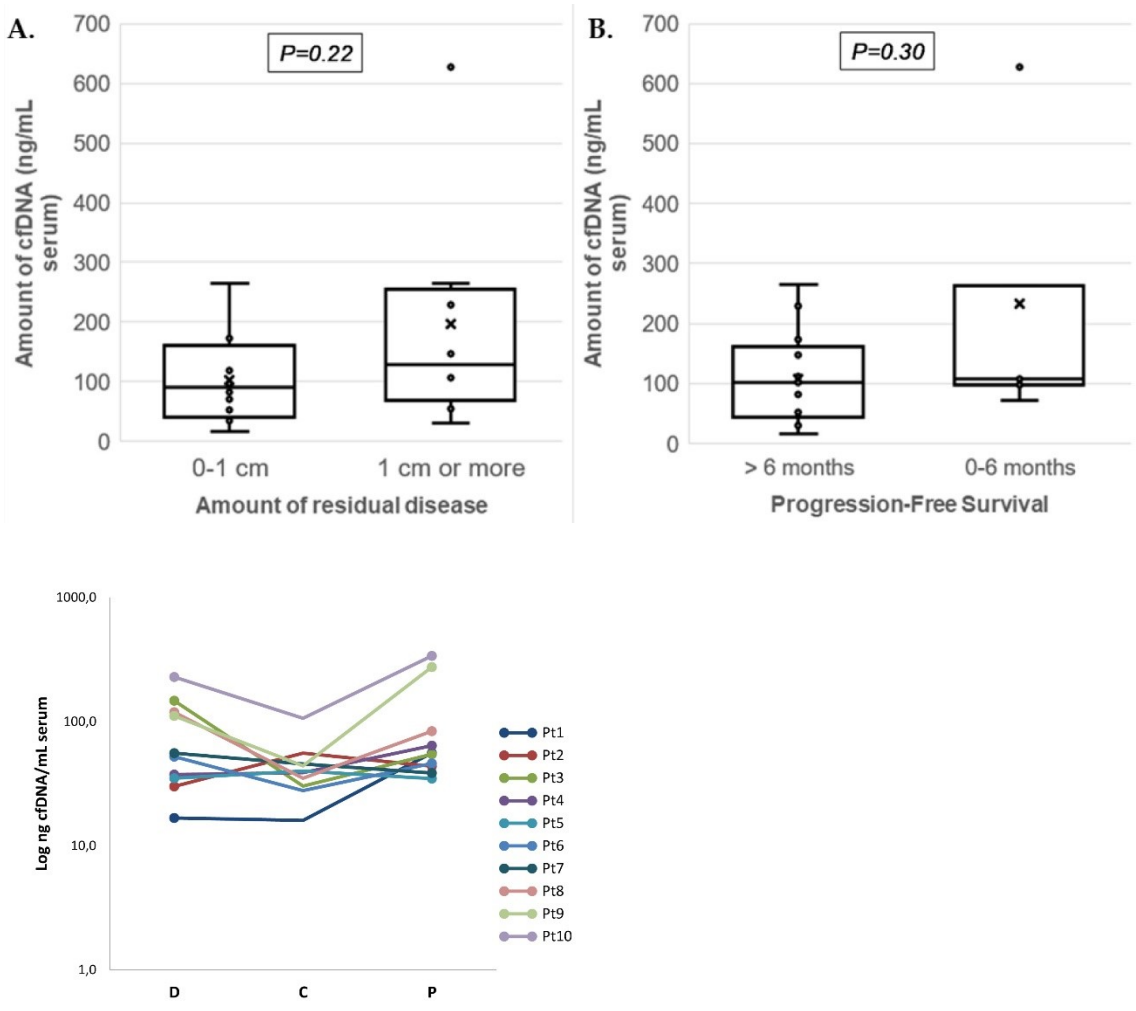


Figure 2. Serum cfDNA yields isolated from patients with advanced ovarian cancer.

Boxplots presenting (**A**) amounts of cfDNA (ng/mL of serum) isolated at diagnosis for patients in relation to residual disease (RD; no or less than 1 cm vs. 1 cm or more) and (**B**) progression-free survival (PFS; ≤6 months and >6 months). The individual measurements are shown as dots, the mean by the cross (x), and

median as horizontal line within the box. The cfDNA amounts between the groups of patients were not significantly different. (**C**) Graph showing the Log cfDNA concentration (ng/mL of serum) isolated from 10 patients at three different time points. Data points correspond to total cfDNA yields per mL serum for each patient (Pt) at the three different time points from serum collection: At diagnosis (D), after chemotherapy (C), and at disease progression (P).

Serum *TP53* Mutation Detection at Diagnosis

In 9/17 patients (53%) with a *TP53* missense mutation in tumor tissue, the mutation was also identified in serum cfDNA at diagnosis (**Table 3; Figure S1**). These mutations were detected in 4/8 patients (50%) by Ampliseq NGS and in 4/6 patients (67%) by digital PCR for workflow I, and by Oncomine NGS in 5/7 patients (71%) for workflow II. dPCR was not performed at diagnosis for patients examined in the latest group due to the low amount of cfDNA available. Overall, *TP53* mutations in serum derived cfDNA at diagnosis were significantly more observed in patients with FIGO stage IV disease ($p = 0.024$; **Table 4**) but not related to other parameters including cfDNA yields or tumor tissue *TP53* mutation frequencies (**Table 4**). Although not statistically significant, the number of patients with mutations detected in serum at diagnosis was higher in patients with non-optimal debulking surgery (4/6 patients (67%)) and disease progression within six months (3/4 patients (75%)) compared to patients with optimal debulking (5/11 patients (45%)) or progression after six months (6/13 patients (46%)). These exploratory findings should be verified in a larger set of patients.

Table 4. Comparison clinicopathological characteristics of high-grade serous ovarian carcinomas (HGSOC) patients with and without serum tumor-specific *TP53* mutation at diagnosis.

HGSOC Patients Serum cfDNA at Diagnosis			
	Without Tumor-Specific *TP53* Mutation	**With Tumor-Specific *TP53* mutation**	***p*-value**
Number of patients	8	9	
Average TP53 mutation allele frequency (MAF in %):			
in tumor tissue	58%	62%	0.819
in cfDNA	0%	7%	0.115
NGS workflow (N):			
I	4	4	0.819
II	4	5	
Age at diagnosis (average):	57	55	0.703
FIGO Stage (N):			
IC	1	0	0.024
IIIC	6	2	
IV	1	7	
Debulking surgery (N):			
PDS	4	3	0.614
IDS	4	5	
Residual Disease (N):			
0–1 cm (optimal debulking)	6	5	0.402
1 cm or more (non-optimal debulking)	2	5	
Progression-Free Survival (PFS):			
0–6 months (n)	1	3	0.312
>6 months (n)	7	6	
average cfDNA yield (ng/mL serum):			
at diagnosis	129	95	0.363
after chemotherapy	31	45	0.106
at progression	117	45	0.268
TP53 mutation at progression measured by dPCR (N):			
Yes	1	4	0.079
No	1	0	
Other gene mutations detected in cfDNA at diagnosis:			
PIK3CA p.H1047R		1.30%	
ESR1 p.R394S	0.30%		
TP53 p.T253I *	0.10%		

PDS = primary Debulking surgery, IDS = interval Debulking surgery, NGS = next generation sequencing, dPCR = digital PCR, N = number of patients, * = *TP53* p.T253I was identified in serum but not in tumor tissue.

Monitoring *TP53* Mutations Over Time

AmpliSeq NGS and dPCR were applied in workflow I to detect *TP53* mutations in sera taken during chemotherapy and at disease progression (**Figure S2**). Both methods were unable to identify mutations during chemotherapy (not shown) but mutations were detected at disease progression in 3/8 patients (38%) by Ampliseq and in 5/8 patients (63%) by dPCR. The cfDNA *TP53* allele frequencies were lower at progression compared to baseline at diagnosis for all patients except patient 1 (**Table 3**). Interestingly, one patient with a *TP53* mutation at progression had this mutation not detected at diagnosis (**Table 3**). Further longitudinal monitoring of cfDNA was performed for patient 5 by the evaluation of CA125 levels and the dPCR monitoring of *TP53* p.K132R in cfDNA derived from additional sera. Low CA125 levels were measured between 5 and 10 months, whereas cfDNA levels increased upon disease progression (**Figure 3**).

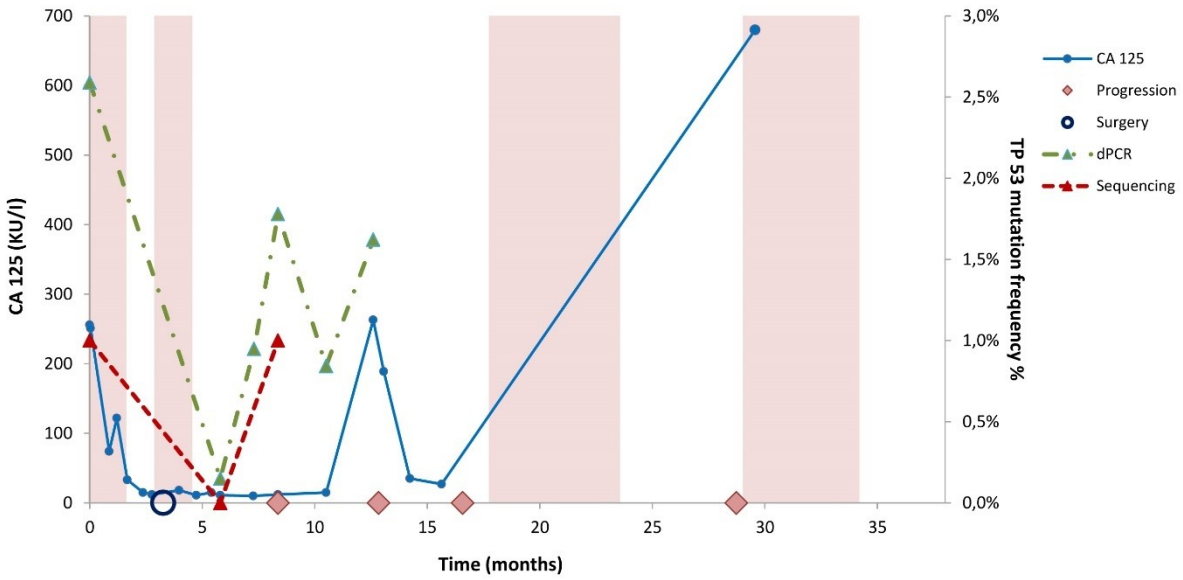


Figure 3. Monitoring cfDNA and CA125 levels over time in patient 5.

Disease monitoring by CA125 levels and *TP53* mutation (p.K132R) levels determined by NGS and digital PCR in patient 5. The colored boxes indicate time on treatment with chemotherapy. The graphs show the change in CA 125 (KU/l) levels and *TP53* mutant allele frequencies (VAF %) in serial serum samples. The somatic mutations were measured using dPCR and Ion Torrent Sequencing. The timeline (in months) is indicated on the x-axis, the allele frequency of the identified mutations is represented on the right y-axis, while the CA 125 level is indicated on the left y-axis. The recolored boxes depict the times on treatment with chemotherapy. Surgery is indicated with a blue circle while clinical observed disease progression is depicted with a pink rhombus.

cfDNA Workflow Comparison

Next to assay detection sensitivity, setup and costs are also important parameters to define an optimal workflow for monitoring *TP53* mutations in blood for (routine) cfDNA evaluation. First, Oncomine-NGS and dPCR (67%) were more sensitive assays than Ampliseq-NGS (50%) for detection of *TP53* mutations in baseline cfDNA. The workflows (**Table 2**) differed in setup to evaluate TP53 only by NGS or combined with dPCR (workflow I) or multiple genes including TP53 by NGS only (workflow II). Secondly, the setup is important, since dPCR is only applicable when the patient-specific *TP53* mutation is known from tumor tissue by Ampliseq-NGS or from cfDNA by Oncomine-NGS. Moreover, NGS enables the detection of additional mutations for *TP53* (workflow I) or for other genes as well (workflow II, **Table 3**), which might be acquired over time in longitudinal cfDNAs. Finally, current estimated costs differ between the two workflows and the applied setup. NGS only will be more expensive than tissue DNA NGS combined with dPCR of cfDNA (**Table 2**). For example, analyses of three longitudinal cfDNAs will cost €450–€750 by Ampliseq-NGS, €1050–€1350 by Oncomine NGS, whereas dPCR costs €460–€490 including €400 for designing the patient-specific *TP53* mutation assay. The differences in costs between NGS and dPCR will increase much more when more longitudinal cfDNAs are evaluated. Workflow I with dPCR is only more cost-effective than NGS when at least 2 cfDNA samples are evaluated.

Discussion

In the current study, we investigated two NGS workflows to detect *TP53* mutations in tissue and serum cfDNA derived from patients with advanced stage serous ovarian cancer. Our aim was to establish the best method and to study whether *TP53* mutations can be used as a tumor biomarker in liquid biopsies. Patient-specific *TP53* missense mutations were identified by targeted NGS in tumor tissue and subsequently analyzed by NGS alone or in combination with dPCR in serum cfDNA taken at different timepoints.

The tumor suppressor *TP53* gene is mainly mutated in exons 4–9, encoding for the DNA-binding domain of the protein [21]. It has been demonstrated that the aberrant protein is able to influence tumor progression toward migration, invasion, and metastasis in different tumor types [22–26]. In our study, 85% of patients harbored non-functional *TP53* alterations in tumor tissue, which is in keeping with previous studies reporting *TP53* mutations in more than 80% of HGSOC

[27–29]. It has been shown by Kang et al. that HGSOC patients carrying a gain-of-function mutant p53 frequently develop resistance against platinum-based treatment as well as being more prone to develop distant metastasis [30]. Therefore, it seems that *TP53* mutations might play a key role in the tumorigenesis of HGSOC.

Levels of cfDNA can vary widely in cancer patients as reported by Fleischhacker and Schmidt [31,32]. These authors reviewed 34 different studies and found that, although the cfDNA concentration in cancer patients is usually much higher than healthy controls, the amount varies widely and is often below 100 ng/mL. Our results are in line with these studies as the cfDNA yields measured at diagnosis were widely variable, although in 55% of cases, the yield was above 100 ng/mL. Yields of cfDNA in blood taken after chemotherapy and progression were, however, lower than 100 ng/mL in almost all cases.

Previous studies have demonstrated a correlation between tumor burden and cfDNA yields. It was shown that highly proliferative lesions release more cfDNA [33,34]. Other studies demonstrated that cfDNA amounts correlated with cancer survival rates [35], and showed its diagnostic value in different tumor types, such as lung cancer [36]. In contrast, we observed in our patient subset no significant differences in serum cfDNA amounts at diagnosis and over time in relation to clinical disease parameters.

Since cfDNA from tumor cells is released in the blood, the cfDNA mutational status may reflect the genetic characteristics of the primary or metastatic lesion [37,38]. Previous studies by Diaz and Bardelli demonstrated that mutations present in tumor tissue are virtually the same as those present in the cfDNA fraction [39,40]. Tumor-specific mutations in cfDNA could therefore act as prognostic and/or predictive biomarkers for cancer patients. In our study, we were able to detect the missense TP53 point mutation present in the primary tumor in cfDNA in 67% of patients. These results are in agreement with previous studies showing that point mutations in TP53 can be also measured in the serum DNA of patients with ovarian cancer [1]. In general, these results highlight the potential of cfDNA as diagnostic tool for ovarian cancer [5,12]. Our small cohort of HGSOC patients with FIGO stage IV disease showed a relation between levels of *TP53* mutations detected in serum cfDNA at diagnosis but less with residual disease or disease progression.

Due to the small amounts of cfDNA and cfDNA available in blood, especially from retrospective archived serum, research is currently focused on the development of new strategies to quantify and characterize cfDNA. Sensitivity and specificity are the main challenges for detecting cancer-

specific alterations in cfDNA. Recent advances in NGS [41–43] and PCR protocols have allowed the quantitative detection of mutations with a sensitivity below 0.001% [44]. Compared to conventional PCR, dPCR is a reliable method and easy to set up. Using this method, it is possible to quantify cfDNA without external references and with a higher sensitivity, precision, efficiency, and reproducibility [40]. In contrast, NGS allows identification of novel genetic or epigenetic mutations. Conventional NGS methods, however, are not as sensitive as dPCR methods and mutations could be missed, particularly when the total number of reads is low. Therefore, researchers have combined the use of NGS and dPCR protocols for liquid biopsy. When comparing the cost and time required for the different techniques, dPCR workflow is much cheaper because it needs less consumables and turnaround time to monitor a specific mutation in follow-up studies compared to NGS protocols. Moreover, dPCR enables accurate quantification of mutant DNA within vast amounts of wild-type DNA, i.e., low mutant allele frequencies (>0.1%), and is often used for the independent validation of NGS results.

Overall, Oncomine NGS in workflow II and dPCR in workflow I enabled detection of *TP53* mutations below 1% allele frequencies and were more sensitive than the conventional Ampliseq NGS. This resulted in the detection of mutations in 67% of patients by dPCR and Oncomine NGS compared to only 50% of patients by Ampliseq NGS. Moreover, to trace *TP53* mutations for disease monitoring in multiple longitudinal serum derived cfDNA, we showed that dPCR is better compared to Oncomine NGS, due to lower amounts for cfDNA needed and lower overall costs. However, a proper direct comparison of mutation detection sensitivity of each platform was not possible due to the limitations of our study design. Summarizing, study restrictions were the limited amounts of archived serum available (>1mL), low cfDNA yields, and different (recommended) cfDNA input amounts for each NGS platform. To accomplish an accurate comparison of detection sensitivities, equal cfDNA input amounts should be evaluated by both NGS platforms in future studies. We also compared the cost-effectiveness of the two workflows, which is of relevance when introducing the methodology to routine diagnostic of HGSOC.

To date, several clinical studies have tried to link *TP53* mutations with patient survival or the development of chemoresistance [45]. However, the conclusions of these studies are often contradictory due to the unselective classification of all *TP53* mutations and the single use of immunohistochemistry to determine the *TP53* mutational status. Recently, it has been proposed that *TP53* mutations could be used as biomarkers to predict patient response to chemotherapy. In line with these studies, we were unable to identify *TP53* mutations at chemotherapy but

detected these mutations at disease progression for 5/6 patients. Parkinson et al. [14] reported that cfDNA *TP53* mutant allele fraction and volumetric measurements were correlated in HGSOC patients, particularly in a subset of patients without ascites. Moreover, almost all subjects with disease volume larger than 32 cm3 showed higher cfDNA copies. A rapid response to chemotherapy was more closely related to cfDNA than to CA-125. These results strongly suggest that cfDNA has the potential to be a highly specific early molecular response marker in HGSOC [14,40].

Conclusions

In conclusion, detection of tumor-specific *TP53* missense mutations in minute amounts of archived serum-derived cfDNA from HGSOC patients is enabled by dPCR or NGS. Our exploratory finding that *TP53* mutations present at diagnosis became undetectable in cfDNA after chemotherapy but re-appeared at disease progression highlights the potential role of *TP53* missense mutations as a biomarker for clinical disease monitoring in ovarian cancer. However, detection sensitivities of NGS platforms need to be validated further in a larger study.

References

1. Giannopoulou, L., S. Kasimir-Bauer, and E.S. Lianidou, *Liquid biopsy in ovarian cancer: recent advances on circulating tumor cells and circulating tumor DNA.* Clin Chem Lab Med, 2018. **56**(2): p. 186-197.
2. Rauh-Hain, J.A., et al., *Ovarian cancer screening and early detection in the general population.* Rev Obstet Gynecol, 2011. **4**(1): p. 15-21.
3. Vitale, S.R., et al., *An Optimized Workflow to Evaluate Estrogen Receptor Gene Mutations in Small Amounts of Cell-Free DNA.* J Mol Diagn, 2019. **21**(1): p. 123-137.
4. Tissot, C., et al., *Circulating free DNA concentration is an independent prognostic biomarker in lung cancer.* Eur Respir J, 2015. **46**(6): p. 1773-80.
5. Dobrzycka, B., et al., *Circulating free DNA and p53 antibodies in plasma of patients with ovarian epithelial cancers.* Ann Oncol, 2011. **22**(5): p. 1133-40.
6. Yu, Z., S. Qin, and H. Wang, *Alter circulating cell-free DNA variables in plasma of ovarian cancer patients.* J Obstet Gynaecol Res, 2019. **45**(11): p. 2237-2242.
7. Kamat, A.A., et al., *Plasma cell-free DNA in ovarian cancer: an independent prognostic biomarker.* Cancer, 2010. **116**(8): p. 1918-25.
8. Capizzi, E., et al., *Quantification of free plasma DNA before and after chemotherapy in patients with advanced epithelial ovarian cancer.* Diagn Mol Pathol, 2008. **17**(1): p. 34-8.
9. Kamat, A.A., et al., *Circulating cell-free DNA: a novel biomarker for response to therapy in ovarian carcinoma.* Cancer Biol Ther, 2006. **5**(10): p. 1369-74.
10. Canzoniero, J.V. and B.H. Park, *Use of cell free DNA in breast oncology.* Biochim Biophys Acta, 2016. **1865**(2): p. 266-74.
11. Cole, A.J., et al., *Assessing mutant p53 in primary high-grade serous ovarian cancer using immunohistochemistry and massively parallel sequencing.* Sci Rep, 2016. **6**: p. 26191.
12. Swisher, E.M., et al., *Tumor-specific p53 sequences in blood and peritoneal fluid of women with epithelial ovarian cancer.* Am J Obstet Gynecol, 2005. **193**(3 Pt 1): p. 662-7.
13. Gormally, E., et al., *TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study.* Cancer Res, 2006. **66**(13): p. 6871-6.
14. Parkinson, C.A., et al., *Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study.* PLoS Med, 2016. **13**(12): p. e1002198.

15. Weerts, M.J.A., et al., *Somatic Tumor Mutations Detected by Targeted Next Generation Sequencing in Minute Amounts of Serum-Derived Cell-Free DNA.* Sci Rep, 2017. **7**(1): p. 2136.
16. Beije, N., et al., *Somatic mutation detection using various targeted detection assays in paired samples of circulating tumor DNA, primary tumor and metastases from patients undergoing resection of colorectal liver metastases.* Mol Oncol, 2016. **10**(10): p. 1575-1584.
17. Jansen, M.P., et al., *Cell-free DNA mutations as biomarkers in breast cancer patients receiving tamoxifen.* Oncotarget, 2016. **7**(28): p. 43412-43418.
18. Schuijer, M. and E.M. Berns, *TP53 and ovarian cancer.* Hum Mutat, 2003. **21**(3): p. 285-91.
19. Kobel, M., et al., *Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma.* J Pathol Clin Res, 2016. **2**(4): p. 247-258.
20. McShane, L.M., et al., *REporting recommendations for tumour MARKer prognostic studies (REMARK).* Br J Cancer, 2005. **93**(4): p. 387-91.
21. Rivlin, N., et al., *Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis.* Genes Cancer, 2011. **2**(4): p. 466-74.
22. Tirro, E., et al., *Chk1 Inhibition Restores Inotuzumab Ozogamicin Citotoxicity in CD22-Positive Cells Expressing Mutant p53.* Front Oncol, 2019. **9**: p. 57.
23. Manzella, L., et al., *New Insights in Thyroid Cancer and p53 Family Proteins.* Int J Mol Sci, 2017. **18**(6).
24. Gasco, M., S. Shami, and T. Crook, *The p53 pathway in breast cancer.* Breast Cancer Res, 2002. **4**(2): p. 70-6.
25. Morton, J.P., et al., *Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer.* Proc Natl Acad Sci U S A, 2010. **107**(1): p. 246-51.
26. Tirro, E., et al., *Molecular Alterations in Thyroid Cancer: From Bench to Clinical Practice.* Genes (Basel), 2019. **10**(9).
27. Ahmed, A.A., et al., *Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary.* J Pathol, 2010. **221**(1): p. 49-56.
28. Kuo, K.T., et al., *Analysis of DNA copy number alterations in ovarian serous tumors identifies new molecular genetic changes in low-grade and high-grade carcinomas.* Cancer Res, 2009. **69**(9): p. 4036-42.

29. Berns, E.M. and D.D. Bowtell, *The changing view of high-grade serous ovarian cancer.* Cancer Res, 2012. **72**(11): p. 2701-4.
30. Kang, H.J., et al., *Clinical relevance of gain-of-function mutations of p53 in high-grade serous ovarian carcinoma.* PLoS One, 2013. **8**(8): p. e72609.
31. Fleischhacker, M. and B. Schmidt, *Circulating nucleic acids (CNAs) and cancer--a survey.* Biochim Biophys Acta, 2007. **1775**(1): p. 181-232.
32. Volik, S., et al., *Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies.* Mol Cancer Res, 2016. **14**(10): p. 898-908.
33. Evan, G.I. and K.H. Vousden, *Proliferation, cell cycle and apoptosis in cancer.* Nature, 2001. **411**(6835): p. 342-8.
34. Beije, N., et al., *Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients.* Mol Oncol, 2018. **12**(1): p. 48-57.
35. Cargnin, S., et al., *Quantitative Analysis of Circulating Cell-Free DNA for Correlation with Lung Cancer Survival: A Systematic Review and Meta-Analysis.* J Thorac Oncol, 2017. **12**(1): p. 43-53.
36. Zhang, R., et al., *Value of quantitative analysis of circulating cell free DNA as a screening tool for lung cancer: a meta-analysis.* Lung Cancer, 2010. **69**(2): p. 225-31.
37. van Dessel, L.F., et al., *High-throughput isolation of circulating tumor DNA: a comparison of automated platforms.* Mol Oncol, 2019. **13**(2): p. 392-402.
38. van Dessel, L.F., et al., *Application of circulating tumor DNA in prospective clinical oncology trials - standardization of preanalytical conditions.* Mol Oncol, 2017. **11**(3): p. 295-304.
39. Diaz, L.A., Jr. and A. Bardelli, *Liquid biopsies: genotyping circulating tumor DNA.* J Clin Oncol, 2014. **32**(6): p. 579-86.
40. Park, Y.R., et al., *Optimization to detect TP53 mutations in circulating cell-free tumor DNA from patients with serous epithelial ovarian cancer.* Obstet Gynecol Sci, 2018. **61**(3): p. 328-336.
41. Kinde, I., et al., *Detection and quantification of rare mutations with massively parallel sequencing.* Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9530-5.
42. Newman, A.M., et al., *An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage.* Nat Med, 2014. **20**(5): p. 548-54.

43. Postel, M., et al., *Droplet-based digital PCR and next generation sequencing for monitoring circulating tumor DNA: a cancer diagnostic perspective.* Expert Review of Molecular Diagnostics, 2017. **18**(1): p. 7-17.

44. Perkins, G., et al., *Droplet-Based Digital PCR: Application in Cancer Research.* Adv Clin Chem, 2017. **79**: p. 43-91.

45. Zhang, M., et al., *TP53 mutation-mediated genomic instability induces the evolution of chemoresistance and recurrence in epithelial ovarian cancer.* Diagn Pathol, 2017. **12**(1): p. 16.

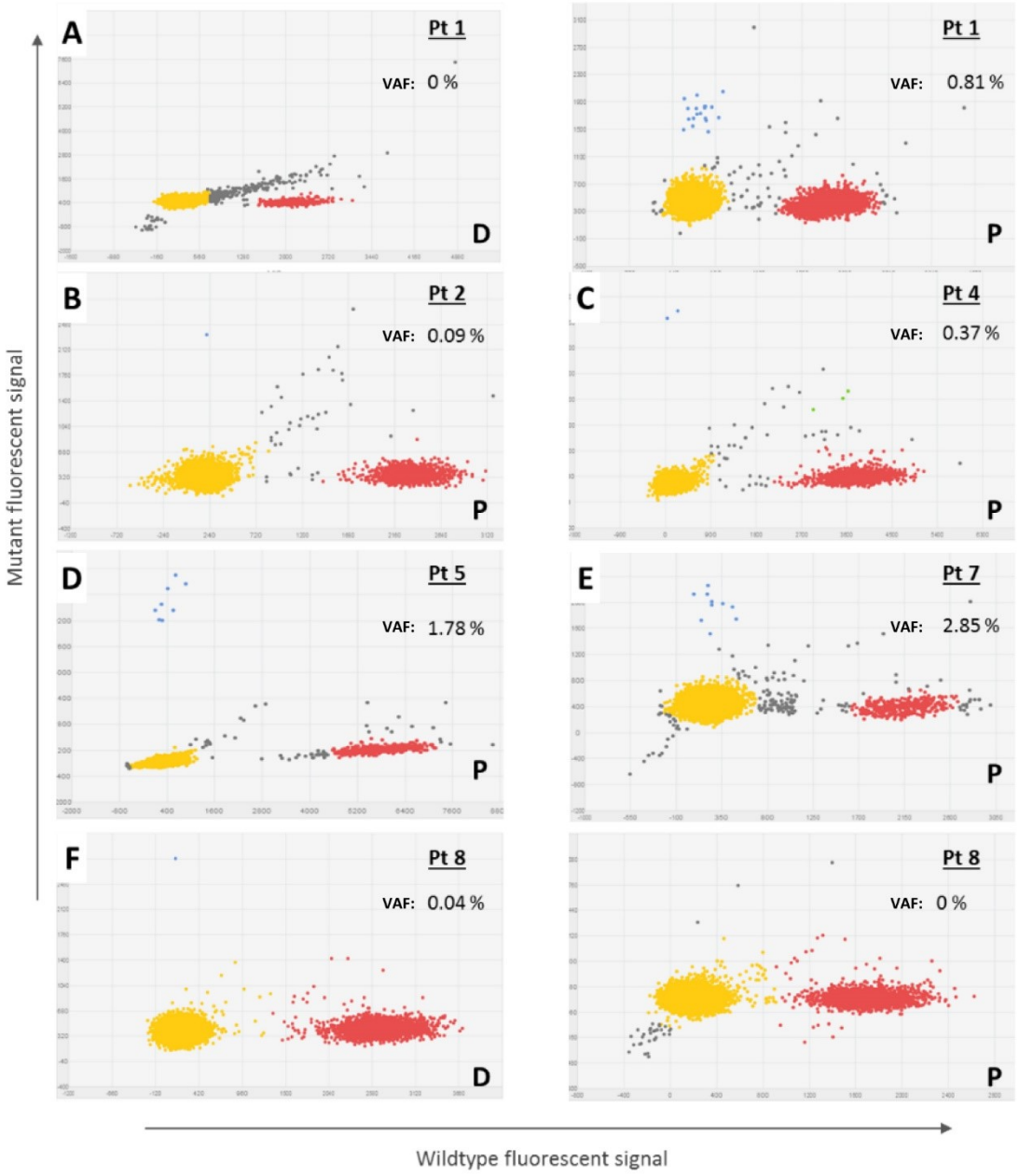


Figure S1. *TP53* mutation analysis by digital PCR in patient serum.

The figure shows dot plots indicating the presence of wildtype (WT) and mutant (MT) copies in cfDNA of **A)** patient 1 analyzed for *TP53*_pY163C. **B)** patient 2 analyzed for *TP53*_pC275Y, **C)** patient 4 analyzed for *TP53*_pR282W, **D)** patient 5 analyzed for TP53_pK132R**, E)** patient 7 analyzed for *TP53*_pY163C and **F)** patient 8 analyzed for *TP53*_pC275Y at diagnosis and/or progressive disease. Blue: wells containing mutant copies, Red: wells containing wildtype copies, Green: wells that contain both wildtype and mutant copies, Yellow empty wells, Grey: undetermined wells, VAF: Variant Allele Frequency. D: diagnosis, P: progression disease

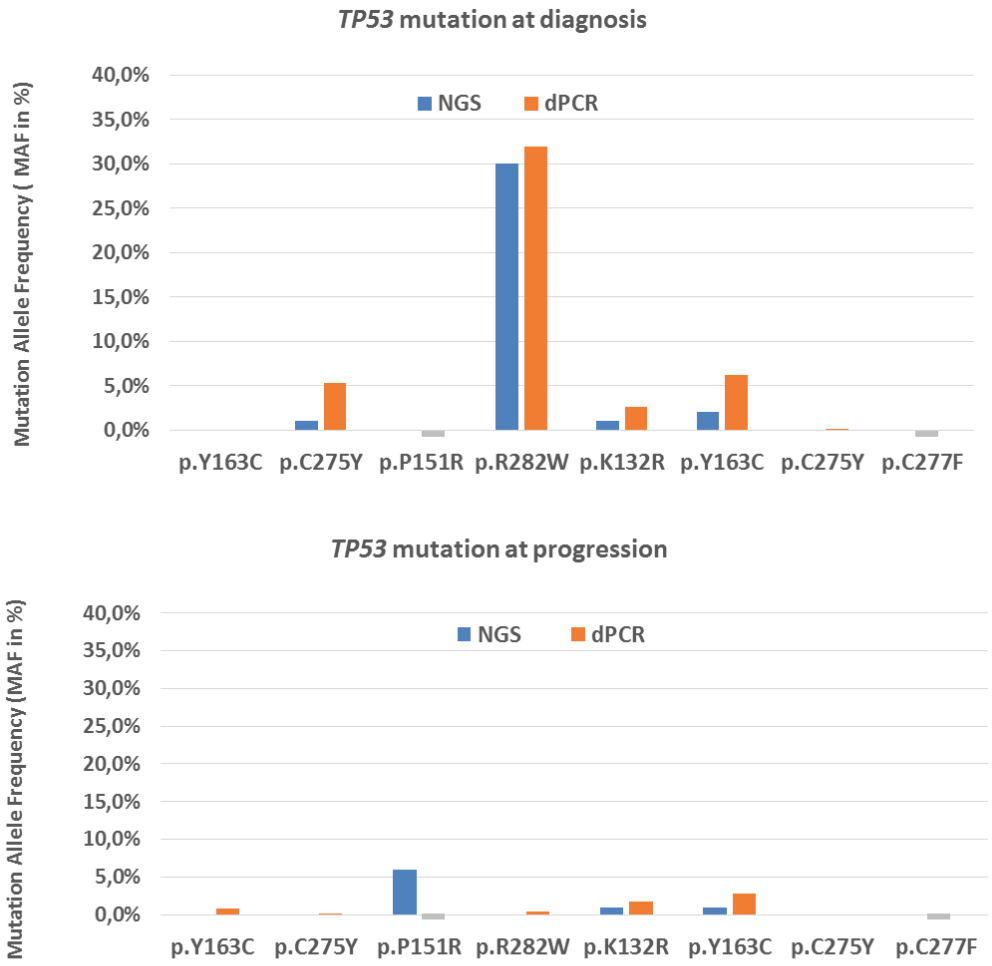


Figure S2. Tumor-specific *TP53* measured by NGS and digital PCR in cfDNA.

Tumor specific *TP53* mutations were measured by Ampliseq NGS (blue bars) and digital PCR (dPCR, orange bars) in cfDNA of archived serum taken at diagnosis, during chemotherapy and at progression. Both NGS and dPCR did not detect any *TP53* mutation in serum during treatment. The TP53_ pP151R and TP53_pC277F mutation were detected by NGS, but dPCR failed (grey bars). The dPCR detected mutations at higher mutation allele frequency than NG and more often at progression (TP53_pY163C, TP53_pC275Y, TP53_pR282W). MAF= Mutation allele frequency; NGS= next generation sequencing, dPCR= digital PCR.

Table S1. Digital PCR SNP genotyping assays used for subset I

Assay ID	Assay Name	Gene	Cosmic ID	Aminoacid Change	Nucleotide Change
AHKA3Z2	TP53_K132R	*TP53*	11582	p.K132R	c.395A>G
AHLJ16A	TP53_P151R	*TP53*	44003	p.P151R	c.452C>G
AHMS0CI	TP53_Y163C	*TP53*	10808	p.Y163C	c.488A>G
AHI15TU	TP53_C275Y	*TP53*	10893	p.C275Y	c.824G>A
AH705J0	TP53_C277F	*TP53*	10749	p.C277F	c.830G>T
AHRSSQL	TP53_R282W	*TP53*	10704	p.R282W	c.844C>T

Table S2. Digital PCR primers and probe sequences

Assay Name	Sequence
TP53_K132R	Fw: 5'-GCAGGTCTTGGCCAGTTG -3' Rev: 5'-GTCTCCTTCCTCTTCCTACAGTACT -3' VIC Probe: 5'-CCCTCAACAAGATGTT -3' FAM Probe: 5'- CCCTCAACAGGATGTT-3'
TP53_P151R	Fw: 5'-TGTGCTGTGACTGCTTGTAGATG -3' Rev: 5'-TGTGCAGCTGTGGGTTGAT -3' VIC Probe: 5'- TCCACACCCCCGCCC-3' FAM Probe: 5'-CACACGCCCGCCC -3'
TP53_Y163C	Fw: 5'-CCTCCGTCATGTGCTGTGA -3' Rev: 5'-GCAGCTGTGGGTTGATTCCA -3' VIC Probe: 5'-CATGGCCATCTACAAGC -3' FAM Probe: 5'-ATGGCCATCTGCAAGC -3'
TP53_C275Y	Fw: 5'- CTGTGCGCCGGTCTCT-3' Rev: 5'-TGGGACGGAACAGCTTTGAG -3' VIC Probe: 5'-TGCGTGTTTGTGCCTG-3' FAM Probe: 5'-TGCGTGTTTATGCCTG -3'
TP53_C277F	Fw: 5'-CTGTGCGCCGGTCTCT-3' Rev: 5'-TGGGACGGAACAGCTTTGAG-3' VIC Probe: 5'-TGTTTGTGCCTGTCCTGG-3' FAM Probe: 5'-TGTTTGTGCCTTTCCTGG-3'

Chapter 5

Detection of tumor-derived extracellular vesicles in plasma from patients with solid cancer

Vitale SR*, Jean A. Helmijr JA*, Gerritsen M, Coban H, van Dessel LF, Beije N, van der Vlugt-Daane M, Vigneri P, Sieuwerts AM, Dits N, van Royen ME, Jenster G, Sleijfer S, Lolkema M, Martens JWM, P.H.M. Jansen MPHM

BMC Cancer 21, 315 (2021)
https://doi.org/10.1186/s12885-021-08007-z

Abstract

Extracellular vesicles (EVs) are actively secreted by cells into body fluids and contain nucleic acids of the cells they originate from. The goal of this study was to detect circulating tumor-derived EVs (ctEVs) by mutant mRNA transcripts (EV-RNA) in plasma of patients with solid cancers and compare the occurrence of ctEVs with circulating tumor DNA (ctDNA) in cell-free DNA (cfDNA).

For this purpose, blood from 20 patients and 15 healthy blood donors (HBDs) was collected in different preservation tubes (EDTA, BCT, CellSave) and processed into plasma within 24 hours from venipuncture. EVs were isolated with the ExoEasy protocol from this plasma and from conditioned medium of 6 cancer cell lines and characterized according to MISEV2018-guidelines. RNA from EVs was isolated with the ExoRNeasy protocol and evaluated for transcript expression levels of 96 genes by RT-qPCR and genotyped by digital PCR.

Our workflow applied on cell lines revealed a high concordance between cellular mRNA and EV-RNA in expression levels as well as variant allele frequencies for *PIK3CA*, *KRAS* and *BRAF*. Plasma CD9-positive EV and GAPDH EV-RNA levels were significantly different between the preservation tubes. The workflow detected only ctEVs with mutant transcripts in plasma of patients with high amounts (>20%) of circulating tumor DNA (ctDNA). Expression profiling showed that the EVs from patients resemble healthy donors more than tumor cell lines supporting that most EVs are derived from healthy tissue.

We provide a workflow for ctEV detection by spin column-based generic isolation of EVs and PCR-based measurement of gene expression and mutant transcripts in EV-RNA derived from cancer patients' blood plasma. This workflow, however, detected tumor-specific mutations in blood less often in EV-RNA than in cfDNA.

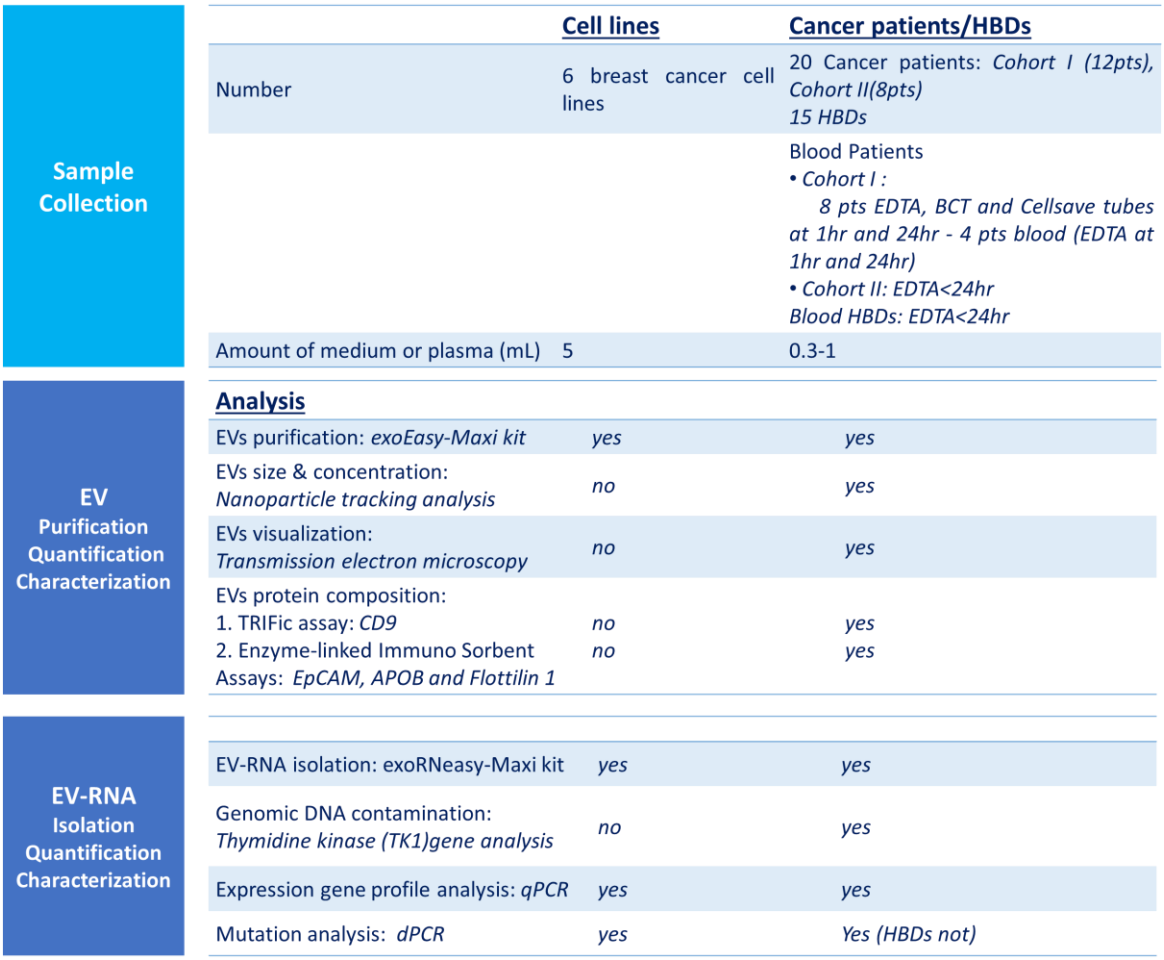
Introduction

Many cell types, including cancer cells [1], release extracellular vesicles (EVs) in various body fluids [2-6]. EVs (size range 40-5000nm) are formed by: 1) Vesicle budding from the cellular membrane (microvesicles), 2) apoptosis (apoptotic bodies) and 3) via the endocytic or the secretory pathway (multi-vesicular bodies) [3, 7-10]. Recent studies showed that EVs are released into the circulation during various pathological processes, including cancer. Circulating tumor-derived EVs (ctEVs) are a small portion of EVs originating from tumor cells which carry their cargo to neighboring cells or distant organs [7, 11, 12]. EVs have heterogeneous membrane compositions and contents [1, 13] and their counts increase over time in blood during disease progression [14]. Most research on EVs was focused on proteomics [15, 16]. EVs contain intact and fragmented mRNA [1, 6, 14], miRNA [7, 17-19], small and long non-coding RNA (ncRNAs), but also tRNAs, and rRNAs [7, 17, 18, 20-22]. Where freely circulating mRNA is prone to rapid degradation outside cells, it is hypothesized that mRNA molecules remain stable within vesicles. Current data favor the hypothesis that mRNAs present in cells do not end up in EVs at random but that only specific mRNA molecules are selectively packaged inside these vesicles [7, 14]. However, before RNA derived from EVs (EV-RNA) can be used as biomarkers for disease detection and for the prediction of prognosis or therapy response in cancer [22], development of reliable detection methods is urgently required. The purpose of this study was to establish such a pipeline to detect ctEVs by establishing a workflow for the isolation and characterization of EVs and EV-RNA. This workflow was firstly tested in cell line models and subsequently applied to analyze EV-RNA isolated from plasma of 20 patients with metastatic cancer. Additionally, we addressed whether our reported pre-analytical conditions established for plasma cfDNA analyses [23] were suitable for isolation and analysis of EVs. Finally, we used this workflow to detect ctEVs by mutant transcripts and evaluate gene expression in both cell line and patient-derived EV-RNA.

Material and Methods

Study Design

Figure 1 shows a detailed overview of the study design.



	Cell lines	Cancer patients/HBDs
Number	6 breast cancer cell lines	20 Cancer patients: *Cohort I (12pts), Cohort II(8pts)* *15 HBDs*
		Blood Patients • *Cohort I :* *8 pts EDTA, BCT and Cellsave tubes at 1hr and 24hr - 4 pts blood (EDTA at 1hr and 24hr)* • *Cohort II: EDTA<24hr* *Blood HBDs: EDTA<24hr*
Amount of medium or plasma (mL)	5	0.3-1
Analysis		
EVs purification: *exoEasy-Maxi kit*	*yes*	*yes*
EVs size & concentration: *Nanoparticle tracking analysis*	*no*	*yes*
EVs visualization: *Transmission electron microscopy*	*no*	*yes*
EVs protein composition: 1. TRIFic assay: *CD9* 2. Enzyme-linked Immuno Sorbent Assays: *EpCAM, APOB and Flottilin 1*	*no* *no*	*yes* *yes*
EV-RNA isolation: exoRNeasy-Maxi kit	*yes*	*yes*
Genomic DNA contamination: *Thymidine kinase (TK1)gene analysis*	*no*	*yes*
Expression gene profile analysis: *qPCR*	*yes*	*yes*
Mutation analysis: *dPCR*	*yes*	*Yes (HBDs not)*

Figure 1. Flow chart of processing samples.
EVs: extra cellular vesiscles; HBDs: healthy blood donors; yes: applied; no: not applied

Cell culture

In this study we used 6 human breast cancer cell lines from the American Tissue Cell Culture (ATCC), i.e. BT20, MDA-MB-231, MDA-MB-361, MDA-MB-435s, MCF7, T47D. Cell lines were cultured at 37oC in humidified air with 5% CO2 in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin, 100 µg/mL streptomycin and 50 µg/mL gentamycin. After cell lines reached 80% confluency, cells were washed with PBS and cultured

RPMI 1640 medium without the addition of FBS (conditioned medium) for 48 hours. Then cells and conditioned medium were collected. The cells were used for DNA and RNA extraction and the conditioned medium was used for Extracellular Vesicles (EV) and EV-RNA purification. Additionally, cell line authenticity was determined by comparing Short Tandem Repeat (STR) analysis using the Powerplex 16 system (Promega, Cat. No: DC6531) with STR marker references from the STR references from the American Type Culture Collection (ATCC).

Sample characteristics and plasma collection

A total of 20 patients with various types of cancer and documented metastases were included in the study, who donated blood within the Erasmus MC Cancer Institute, Rotterdam, Netherlands. Blood was donated in 2015-2016 for Cohort I (12 patients) and in 2017 for cohort II (8 patients). None of the patients received systemic treatment at time of blood draw. Blood from 15 healthy blood donors (HBDs), collected between September 2016 and September 2017, were provided by the Sanquin Blood Bank South-West Region (The Netherlands). The study was approved by the institutional review board of the Erasmus MC Medical Ethical Committee (Erasmus MC ID: MEC-15-616) and conducted in accordance with the Declaration of Helsinki and applicable regulatory requirements. The study was carried out according to the REMARK guidelines and Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (https://www.federa.org/codes-conduct). All patients provided written informed consent before blood collection and data analysis. The characteristics of the patients are summarized in **Table 1**. All patients had a known somatic variant in their primary tumor or metastatic lesion, which was also detected in plasma cfDNA from 9/12 patients (Cohort I) and all 8 patients (Cohort II).

The blood collection methods for Cohort I were previously described [23]. Briefly, blood was collected in three types of vacutainer tubes (EDTA, Cell-Free DNA BCT, CellSave) and processed into plasma at 1 hour, and 24 hours after blood draw. Plasma was obtained from the blood after centrifugation at 1711g for 10 minutes followed by 12000g for 10 minutes, both at room temperature. From the 16 cases of the previous study [23], only 12 patients were eligible in the current study for detection of mutant transcripts in EV-RNA with digital PCR (dPCR) mutation assays. Plasma (range 0.3-1.0mL) processed at 1 hour and 24 hours after blood draw was evaluated for all 12 patients when collected in EDTA and for 8 patients when collected in BCT and CellSave tubes. For 15 HBDs and Cohort II of 8 metastatic cancer patients, blood was collected in

EDTA tubes and processed into plasma within 12-24 hours (< 24h) after venipuncture, by using the above-described protocol. Then, plasma was stored at -80oC in 1 mL aliquots until further processing.

Table 1. Characteristics of the clinical samples included in this study.

	#	Primary tumor	Mutation in primary tumor	*EV-RNA isolation*		*TRIFic CD9 Analysis*	*Target gene analysis*		*Mutation Analysis*	*Variant Allele Frequency*		
				Plasma Input (mL)	Yield[a] (ng)	CD9-EV Europium[a] (count × 10^5)	GAPDH[a] (cp × 10^4)	Target gene[a] (cp × 10^2)	MT|WT cp[a] in EV-RNA	EV-RNA (%)	Primary Tumor (%)	cfDNA[c] (%)
Cohort I	*1*	**Cholangio Carcinoma**	*KRAS* G12D (c.35G>A)	0.8	646	1.1	2.8	7.8	0|600	0	40	0
	2	**Colorectal Carcinoma**	*PIK3CA* H1047R (c.3140A>G)	0.7	908	3.7	0.2	0	0|0	0	38	2.7
	3	**Breast Cancer**	*PIK3CA* H1047L (c.3140A>T)	0.8	301	2.9	4.1	6.7	0|432	0	26	0
	4	**Melanoma**	*BRAF* V600E (c.1799T>A)	0.8	1073	1.9	1.6	2.1	0|395	0	3	1
	5[b]	**Colorectal Carcinoma**	*KRAS* G13D (c.38G>A)	1.3	372	8.4	5.3	10	202|1161	15	60	64.5
	6	**Colorectal Carcinoma**	*KRAS* G12D (c.35G>A)	0.8	250	2.5	0.9	3.1	0|315	0	44	7.3
	8	**Melanoma**	*BRAF* V600E (c.1799T>A)	0.7	820	4.9	6.4	8.8	0|475	0	64	35.7
	9	**Melanoma**	*BRAF* V600E (c.1799T>A)	0.65	205	1.5	2.2	2.6	0|373	0	70	4.2
	13	**NSCLC**	*EGFR* T790M (c.2369C>T)	0.8	473	1.2	1	0.2	0|0	0	17	0.9
	14	**Melanoma**	*BRAF* V600E (c.1799T>A)	0.8	655	0.8	1.5	1.9	0|410	0	56	5.8
	15[b]	**NSCLC**	*EGFR* T790M (c.2369C>T)	1.05	368	4.1	2.7	3	288|938	24	65	26.3
	16	**Melanoma**	*BRAF* V600E (c.1799T>A)	0.8	137	1.2	5.3	6.6	0|756	0	50	0

	17	**NSCLC**	*KRAS* G12C (c.34G>T)	0.9	130	5.6	1.9	5.9	0|673	0	32	0.5
	18	**Melanoma**	*BRAF* V600E (c.1799_1800delinsAA)	0.9	154	3.3	1.2	1.3	0|184	0	50	4.4
	19	**Melanoma**	*BRAF* V600K (c.1798_1799delGTinsAA)	0.6	174	4.3	1.8	3.1	0|556	0	38	2.6
	20	**Colon Carcinoma**	*KRAS* G12D (c.35G>A)	0.9	414	3.1	3.6	1.1	0|2322	0	45	4.5
Cohort II	*21*[b]	**Colon Carcinoma**	*KRAS* G13D (c.38G>A)	0.9	487	4.8	3.6	1.1	34|1521	2.2	40	23.8
	22	**Rectum-carcinoma**	*KRAS* G12V (c.35G>T)	1	202	4.4	1.3	3.5	0|511	0	UNK	0.9
	23	**Colon Carcinoma**	*KRAS* G13D (c.38G>A)	1	154	9.6	3.0	7.7	0|949	0	50	2.3
	24	**Melanoma**	*BRAF* V600K (c.1798_1799delGTinsAA)	0.9	294	3.3	5.0	8.6	0|786	0	55	0.9

[a] per mL plasma

[b] Mutant EV-RNA transcripts detected

[c] cfDNA analyses have previously been performed and described (van Dessel, L. F. et al. Mol Oncol, doi:10.1002/1878-0261.12037 (2017)

VAF: Variant Allele Frequency, EV: Extracellular Vesicle(s), NSCLC: Non-Small Cell Lung Cancer, UNK: Unknown, MT: Mutant, WT: Wild-type, cp: copies

Isolation of Extracellular Vesicles

The exoEasy-Maxi kit (Qiagen, Cat. No: 77064) was used according to the manufacturer's protocol to isolate EVs from 5mL of conditioned culture medium and 0.3-1.0mL of plasma. Briefly, medium and plasma were filtered using a 0.8µm syringe filter (Millipore, Cat. No: SLAA033SS) to remove larger particles such as apoptotic bodies and cell fragments, followed by purification of EVs by mixing 1 volume of sample (filtered medium or plasma) with 1 volume of XBP buffer. Subsequently, this mix was added to an ExoEasy spin column, centrifuged for 1 minute at 500g and 4^{o}C. The membrane bound EVs were obtained by eluting with 400 µL XE buffer and centrifugation for 5 minutes at 500g and 4oC. The isolated EVs from HBDs and patients plasma were subsequently quantified and characterized following MISEV2018 guidelines [24]. For this, EVs were evaluated by nanoparticle tracking analysis, immuno-assays and transmission electron microscopy.

Nanoparticle tracking analysis of extracellular vesicles

The sizes and concentration of extracellular vesicles were evaluated by NTA using the NanoSight NS300 (Malvern Panalytical, Malvern, UK), with a blue laser 488nm and sCMOS camera. EV pellets from plasma were diluted 1:1000 (v/v) with Phosphate Buffered Saline (PBS). Each sample was recorded and analyzed for one minute in five replicate measurements by NTA 3.0 software to determine particle concentration and sizes.

Transmission electron microscopy (TEM)

EVs purified from plasma were evaluated by TEM. Plasma EV-pellets were diluted in 10 µL PBS and 10 µL of this was added on to a formvar/carbon-coated 400 mesh copper grid for 7 minutes. Grid staining was performed with Uranyless EM Stain for 1 min (negative stain). Grids were air-dried, and visualized with an TALOS L120C TEM at 120 kV at 11k-45k magnification.

Protein content characterization of extracellular vesicles

The EVs from plasma were characterized for (non-)tissue specific transmembrane protein CD9, and EpCAM (MISEV2018 Category 1), cytosolic protein FLOT1 (Category 2), and non-EV co-isolated structures by apolipoprotein APOB (Category 3). These analyses were performed by

Europium Time-Resolved Immunofluorescence assays (TRIFic) assay for CD9 or by Enzyme-linked Immuno Sorbent Assays (ELISA)) for EpCAM, APOB and FLOT1.

The protein content of EVs were further evaluated with ELISA assays for EpCAM (Abcam, cat# ab264632), APOB (Abcam, cat# ab108807), and FLOT1 (Aviva System Biology, cat# OKEH02189). The ELISA assays were performed according to the protocols of manufacturers using with 50-100 µL of isolated EVs purified with the exoEasy-Maxi kit. In Short, provided standards were diluted as instructed and 100µL of each dilution was pipetted in the provided 96 wells microtiter plate. If less than 100µL of sample was used, the volume was brought up to 100µL with XE-buffer (ExoEasy elution buffer) for EV preparations or PBS for plasma samples. Both XE-buffer and PBS were used as background signal controls. To prevent inter experiment variations all samples were analyzed on the same microtiter plate and time for each Elisa. All ELISA incubation and washing steps were performed with gentle shaking of solutions. Absorbances were measured at 450 nm with a microplate reader (version 5.2, Bio-Rad) software and OD values were corrected for background signal. The generated standard curves were used to calculate the protein concentrations. Additionally, all protein concentrations were normalized based on the plasma volume used for EV purification.

In Contrast to the ELISA analysis, The EV marker CD9 was directly analyzed in plasma specimen or conditioned culture medium using TRIFic exosome assays (CD9: Cat. No.: EX101, Cell Guidance Systems) following the manufactures guidelines and as previously reported [25]. Briefly, streptavidin-coated plates were incubated with biotinylated CD9 for 1 hour at room temperature. Supernatant was removed and plates were washed with buffer (Kaivogen Oy, Turku, Finland, Cat. No: 42-01) in an automated plate washer (TECAN Columbus). Subsequently, 10µL of filtered plasma or culture medium and 90µL PBS were transferred to the wells, incubated for 1 hour and washed and incubated with 100µL Europium-labeled CD9 antibodies for 1 hour, all at room temperature. After another wash step, 100µL enhancement solution was added and incubated for 15min at room temperature. Europium time-resolved fluorescence was subsequently measured at 615nm wavelength by a Wallace Victor2 fluorometer (Perkin Elmer, Cat. No: 1420-020).

Cellular and EV-RNA isolation

RNA was extracted from cells collected after incubation on conditioned medium was using the RNeasy mini kit (Qiagen, Cat. No: 74104) according to the manufacturer's protocol. In Summary,

after filtering the medium or plasma and binding EVs to the exoEasy spin Column, EVs bound to the silica membrane were lysed using 700µL of Qiazol (Qiagen, Cat. No: 79306) and the QIAzol RNA mix was collected by centrifugation for 5min at 4^{o}C and 500g. Then samples were thoroughly mixed with chloroform and followed by centrifugation at 12,000g and 4^{o}C. After multiple washing steps, the purified (EV-)RNA was eluted in 20 µL RNase-free water and quantified with the Thermo Scientific NanoDrop 2000 Spectrophotometer.

DNase treatment and cDNA generation

Prior to cDNA generation, 50-400 ng (EV)-RNA was pretreated with 1 Unit RNase-free DNase I (New England Biolabs, Cat. No: M0303S) at 37oC for 10min to remove contaminating DNA. DNase was inactivated by 1 µL EDTA (50 mM) at 75oC for 10min. The resulting 10µL of DNase treated sample was used to generate cDNA with the SuperScript VILO cDNA Synthesis Kit (Thermo Fisher, Cat. No: 11754250) according to the manufacturer's protocol. After cDNA generation, samples were treated with 2 Units of Ambion RNase H (Thermo Fisher, Cat. No: AM2293) and incubated at 37^{o}C for 30min to remove any residual RNA.

A 136bp fragment located in intron 2 of the Thymidine kinase (TK1) gene was analyzed in our ER-RNA samples after DNase treatment and cDNA generation to demonstrate successful removal of DNA contamination, using the following primers and probes:

Forward primer: 5'-CTCTGGGAACAACTCTGGGATGAGG-3'; Reverse primer: 5'-ACTCAGGTGGTCCCAGGAAGTGTGG-3' and labeled MGB probe sequence: 5'-FAM-GAAGGCAG-3'. The analysis was performed on the Quant 3D Studio digital PCR system (see below).

Gene expression profile analysis

Simultaneous expression analysis of 96 genes previously reported by us [26, 27] was performed in duplicate on total cellular RNA and on EV-RNA from 5 cell lines, and on EV-RNA of 6 patients. The selected genes were more abundant expressed in tumor cells than in white blood cells [27]. After cDNA generation, linear multiplex pre-amplification was applied using 96 target specific Taqman assays and TaqMan PreAmp according to the manufacturer's instructions. Next, pre-amplified cDNA preparations were analyzed in a Mx3000P Real-Time PCR System (Agilent, Amsterdam, The Netherlands), with individual TaqMan Gene Expression Assays and TaqMan

Universal PCR Master Mix, no AmpErase UNG (Thermo Fisher, Cat. No: 4324018). Levels of the reference genes HMBS, HPRT1 and GUSB were used to control sample loading and RNA quality [26, 27]. Gene expression levels were quantified using the delta quantification cycle threshold (ΔCt) method i.e. is the difference between the average Ct of the reference genes minus the Ct of the target gene. EV-RNA expression profiles from cell lines were compared with their cellular mRNA expression profiles to characterize genes with enriched expression in EVs. Patient plasma EV-RNA was also evaluated and was from plasma without circulating tumor DNA (ctDNA) nor circulating tumor EVs (ctEVs) (P1, P3), plasma with only ctDNA (P6, P13) and plasma with both ctDNA and ctEVs (P5, P15). Additionally, the EV-RNA expression profiles from 6 patients were compared with those from cell lines and with the median leucocyte mRNA expression profile of 53 HBDs [26, 27], to define tumor cell related gene expression in patient EVs. The mRNA expression profiles were generated for all cell lines except for MDA-MB-435s.

Genomic DNA contamination, mRNA target gene transcript quantification and mutation detection by digital PCR

The variant allele frequencies (VAF) and number of mutant and wildtype (EV-) RNA transcripts were evaluated in both cell line and patient EV-RNA samples for 4 known oncogenes using the QuantStudio 3D digital PCR system (Thermo Fisher, Cat. No: A29738). The used mutation-specific TaqMan assays, summarized in Table S1, were: PIK3CA p.H1047R(c.3140A>G) and p.H1047L(c.3140A>T), KRAS Screening Kit, KRAS p.G12C(c.34G>T), p.G12D(c.35G>A) and p.G12V(c.35G>T), EGFR p.T790M(c.2369C>T),.In addition BRAF was evaluated by a multiplex of the assays BRAF wild-type and p.V600E(c.1799T>A), and custom made exon spanning BRAF p.V600E(c.1799T>A) assay next to BRAF p.V600E(c.1799_1800delinsAA) and p.V600K(c.1798_1799delGTinsAA). Taqman expression assays were used with a FAM labeled MGB-probe for PIK3CA, KRAS, BRAF or EGFR and multiplexed with the VIC labeled MGB-probe for GAPDH (Cat. No: 4326317E). All dPCR assays were performed on the ProFlex 2 x flat PCR System (Thermo Fisher, Cat. No: 4484078) thermal cycler in combination with the QuantStudio 3D Digital PCR Chip Adapter Kit (Thermo Fisher, Cat. No: 4485513) using the following program: 1 cycle of 10min at 96°C, 40 cycles extension/annealing of 2min at 56°C for the Thermo Fisher assays or at 52°C for the Bio-Rad assays, 30sec at 98°C and 1 cycle of 2min at 56°C and terminated at 10°C. After amplification, data were acquired using the Quantstudio 3D dPCR instrument and analyzed with the web-based Quantstudio 3D dPCR Analysis Software version 3.01 (Thermo

Fisher). At least one positive and one no-template control (RNAse-free Water) were used for each assay to determine the thresholds for calling positive mutant and wildtype copies. The software automatically calculated the VAF by dividing the number of mutant copies by the total measured copies (wildtype + mutant). Presence of mutant EV-RNA transcripts was determined for the cell lines BT20, MDA-MB-231 and MDA-MB435s but not for MDA-MB-361 and MCF-7 for which the PIK3CA p.E545K mutation detection assay failed on EV-RNA.

Statistical analysis

Protein levels of EpCAM, APOB, FLOT-1 and CD9 were evaluated in Patient and HBD derived EV's and tested for statistical differences using the two-tailed T-test. P-values lower than 0.05 were considered significant. To identify differentially expressed genes between RNA isolated from EVs and tumor cells, a paired class comparison analysis was performed on the generated 96 genes expression profiles of both cellular RNA and EV-RNA using BRB-ArrayTools version 4.5.0 ("http://linus.nci.nih.gov/BRB-ArrayTools.html"). False discovery rate (FDR) was set at 10% to correct for multiple testing and the significance threshold of the univariate tests was set to P-value <0.05 33. Genes that passed these criteria were considered differentially expressed.

Results

Our study evaluated EV preparations of plasma, pre-analytical conditions on CD9-positive EVs amounts and the downstream analysis of EV-RNA in cell line models, healthy blood donors (HBDs) and two cohorts of patients, and describes a strategy to detect ctEVs in plasma by mutant transcripts and gene expression.

Characterization of EV preparations

The EVs of plasma from solid cancer patients and HBDs were harvested by the ExoEasy protocol and were evaluated for specific characteristics described by MISEV2018 guidelines (**Figure 2**). No significant differences in EV preparations between patients and HBDs were observed for nanoparticle concentration (P=0.25) and EV protein analysis of FLOT1 (P=0.31) but a significant difference was observed for nanoparticle size(P=0.001), and for protein levels of APOB (P<0.02)

and of EpCAM (P<0.001) **(Figures 2 A-E**). TEM demonstrated the presence of EVs as exemplified for a patient preparation (**Figures 2 F, G**).

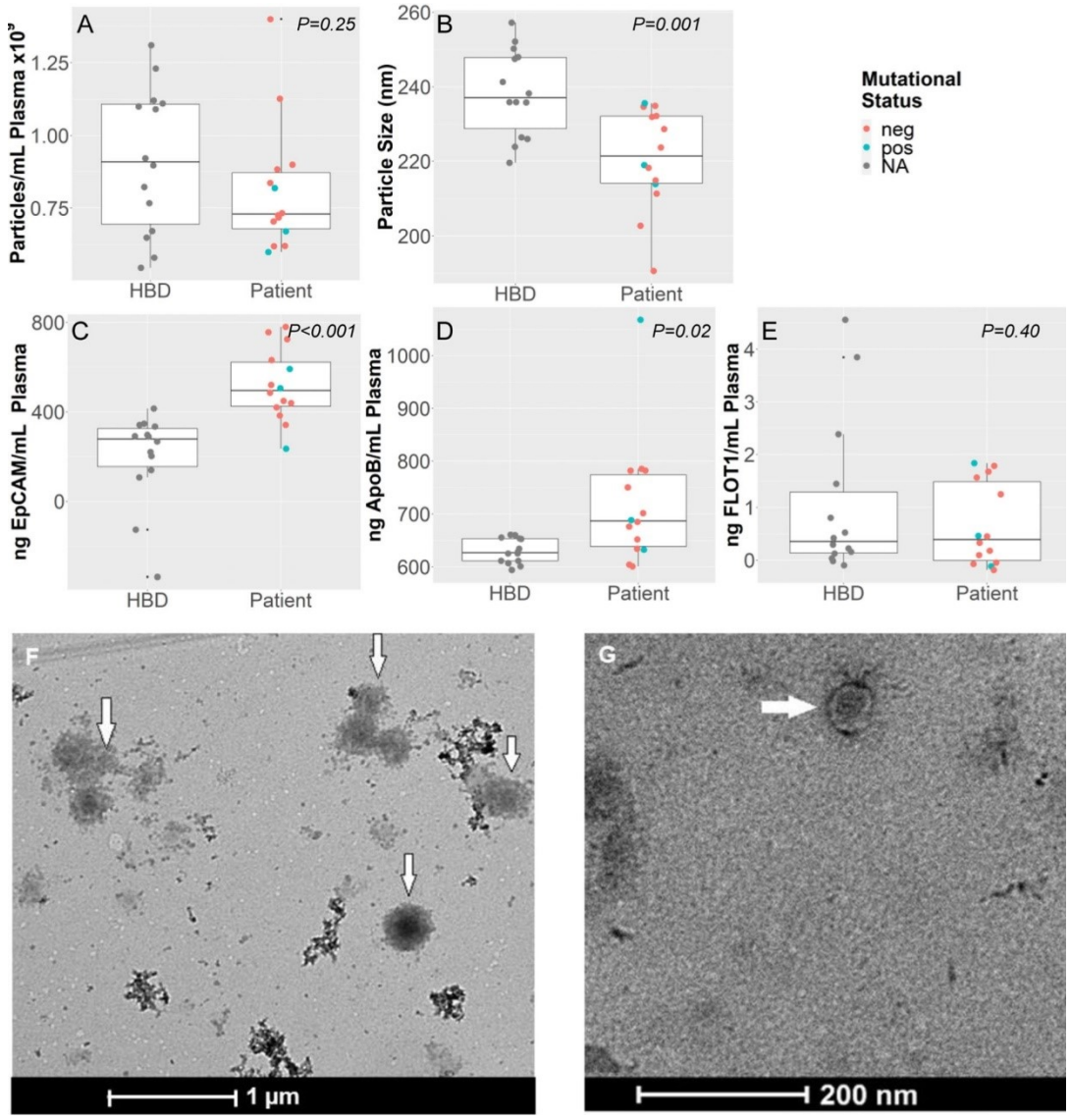


Figure 2. Characterization of EV-preparations.

Plasma EV-preparations from patients and HBDs containing EVs isolated by the Exo(RN)easy kit (Qiagen) were characterized by NTA (A,B), immune-assays (C-E), and TEM (F, G)). Figures A-B demonstrate differences in nanoparticle concentration (A) and size (B) between healthy blood donors (HBD) and patients (PT). Figures C-E illustrate the EV protein content measured for EpCAM, APOB and FLOT-1. EV-preparation of patients have more EpCAM but less APOB compared to HBDs. The EV cytosolic protein FLOT-1 was comparable between both EV-preparations. Figures F and G shows EVs, indicated by arrows, as visualized by TEM from EV-preparations of patient 15 plasma. Pictures were taken at 11k (F) and 45k (G) magnification.

Pre-analytical conditions and amounts of CD9-positive EVs

Extracellular vesicles derived from conditioned cell culture medium and from plasma of 15 HBDs and 8 patients were quantified by the TRIFic CD9 assay. The CD9 levels in conditioned medium from all cell lines were well above (≥1.4×) the levels measured in culture medium alone (**Figure S1**). Fetal Bovine Serum (FBS) gave no signal, confirming the CD9 antibody specificity for human EVs (data not shown). The plasma CD9 levels in cohort I of 8 patients were compared between blood collected in EDTA, BCT or CellSave tubes and processed into plasma at 1 hour or at 24 hours after blood draw. For samples processed at 1 hour, CD9 levels were significantly higher for BCT and CellSave tubes (both $P<0.008$) compared to standard EDTA tubes (**Figure 3A**). All tube types had higher CD9 levels in plasma processed at 24 hours compared to plasma processed at 1 hour ($P<0.05$) (**Figure 3A**). Higher CD9 levels were also observed in patients of cohort II compared to HBDs in blood of EDTA tubes processed into plasma within 24 hours (<24h), however, not significantly different (**Figure 3B**). Finally, CD9 levels in EDTA tubes gradually increased with time to process blood into plasma after venipuncture ($P<0.05$) (**Figure S2A**).

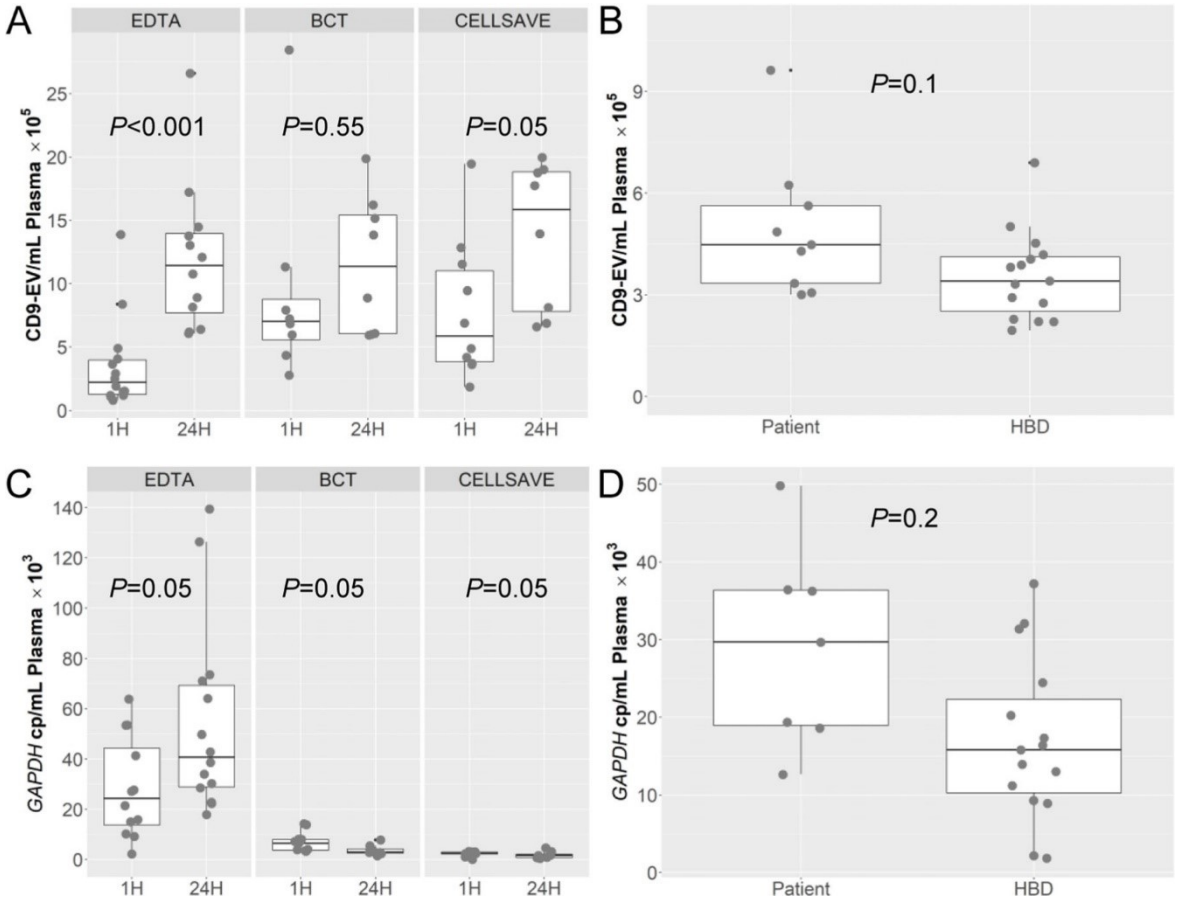


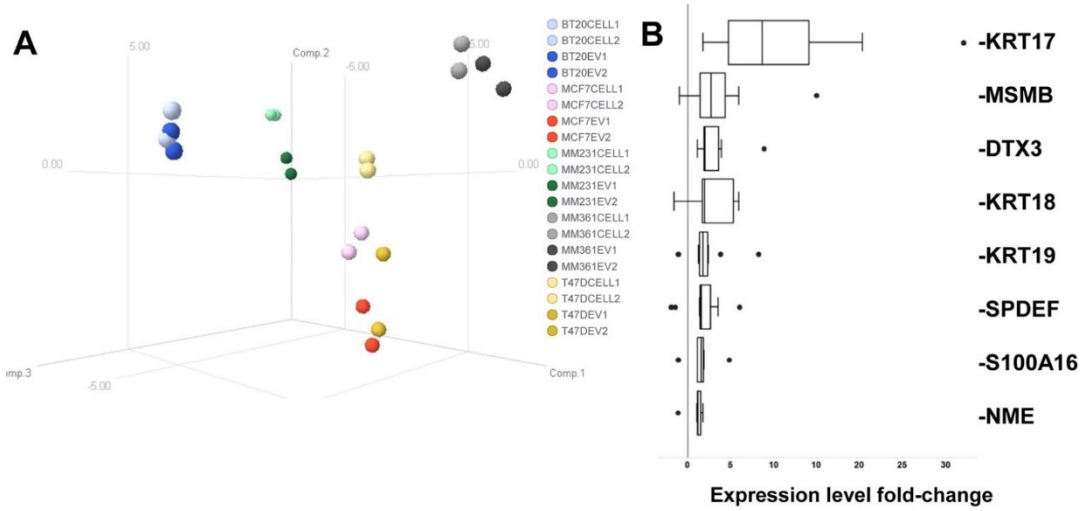
Figure 3. Analysis of CD9 and *GAPDH* expression in EVs from plasma of cancer patients.

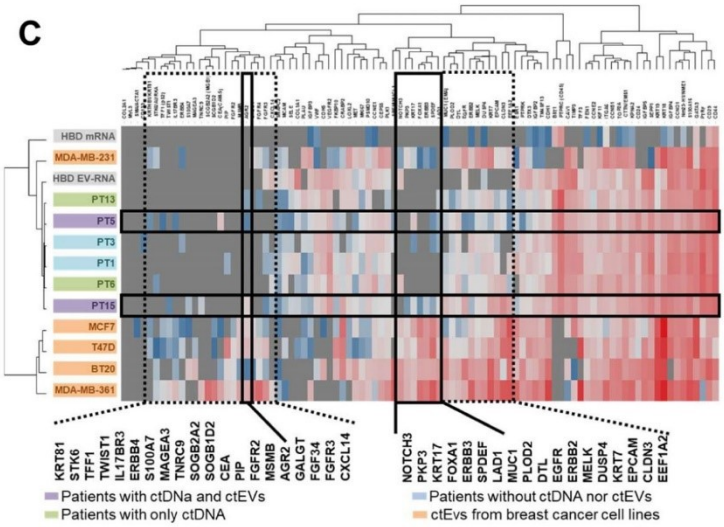
CD9-EV levels on EVs were measured by TRiFIC and *GAPDH* transcripts in EV-RNA were determined by digital PCR. The boxplots shows for 8 patients of cohort 1 in A) CD9 levels per mL plasma and C) *GAPDH* copies/mL plasma measured in plasma collected in different tubes (EDTA, BCT and CellSave) and processed at two time points (1 hour and 24 hours). For the second cohort of 8 patients and 15 Healthy Blood Donor (HBDs) are boxplots presented in B) CD9 levels per mL plasma and D) *GAPDH* copies/mL plasma both collected in EDTA tubes and processed within 24 hours (< 24 h) from the blood draw. CD9 measurements were performed in duplicate**.**

Comparison of EV-RNA and matched tumor cell mRNA gene expression profiles

To investigate whether the transcriptome was equivalent between EVs retrieved from conditioned medium and the cells they originated from, we compared the expression of 96 genes in 5 breast cancer cell lines and their respective EVs. The cell lines included both basal (MDA-MB-231, BT20) and luminal (MCF7, T47D, MDA-MB-361) molecular breast cancer subtypes. The 96 gene expression profiles of EV-RNA and matched parental cell line mRNA correlated highly (RPearson >0.85, P<0.05; Table S2), and grouped together after hierarchical clustering (**Figure S3**)

and principal component analysis (PCA) for all cell lines except T47D (**Figure 4A**). Although expression profiles between cell lines and matched EVs were highly comparable, paired class comparison analyses revealed that the expression levels of 38 genes (Figure S4 and Table S3) were different between EVs and tumor cells (P<0.05). Specifically, 8 genes (DTX3, KRT17 KRT18, KRT19, MSMB, NME1, S100A16, SPDEF) were enriched in EVs (**Figure 4B**).





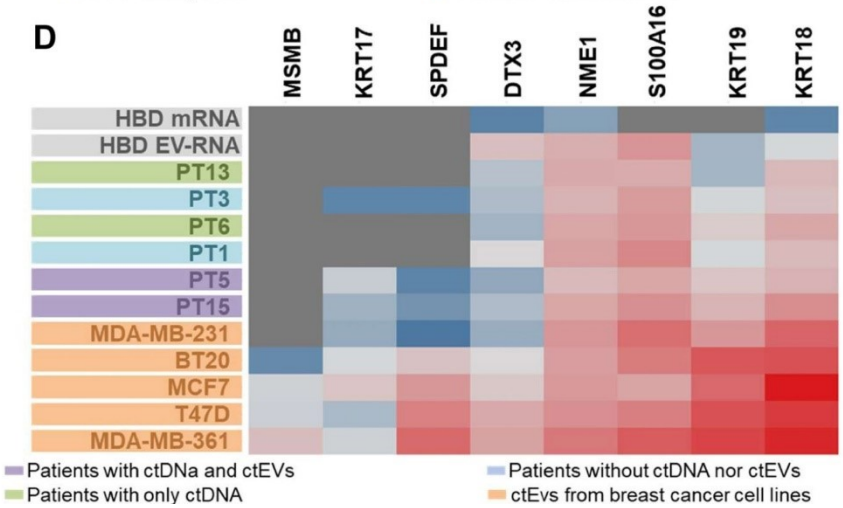


Figure 4. Gene expression profiling of EV-RNA.

Expression of 96 genes was evaluated by RT-qPCR in EV-RNA and cellular mRNA from cancer cell lines and HBDs and EV-RNA from patients. **A**) 3D Principal Component analysis plot of the 96 gene expression profiles in EVs and matching tumor cells from 5 breast cancer cell lines. Each dot represents a sample. Axes show the first three principal components. Dark colored dots indicate the EV-RNA samples and the lighter colors the cellular mRNA. **B**) Enriched genes in EVs are indicated on the Y-axis while the X-axis shows the expression fold-change in EV-RNAs compared to their matched cellular mRNA from 5 breast cancer cell lines. Each boxplot consists of data from duplicate analysis of EV-RNA and tumor cell mRNA. **C**) Hierarchical cluster analyses of EV-RNA expression profiles from patients, HBD and breast cancer cell lines compared to median HBD leucocyte mRNA expression. Clustering is shown for 96 genes (upper plot) and for the 8 genes enriched in tumor cell line EVs (lower plot; see also figure 2B). The boxes in the upper plot indicate genes upregulated in EVs from tumor cells compared to HBDs (dashed black boxes), and tumor cell related genes (solid black boxes) expressed in cell lines and in patients with ctEVs and ctDNA (PT5 and 15). Expression levels of target genes were compared to the average of reference genes *HMBS*, *HPRT1* and *GUSB* levels; grey color indicates no expression, blue color is below reference gene level, and red color is above reference gene level.

Workflow for gene expression and mutant transcript analysis in EV-RNA

Next, we tested our workflow for mutant and wildtype transcript detection in cellular mRNA and EV-RNA using the QuantStudio 3D digital PCR and Taqman mutation assays for KRAS p.G12D and p.G13D, PIK3CA p.H1047R, EGFR p.T790M and BRAF p.V600E. We designed an exon spanning BRAF p.V600E assay for RNA templates only. All other mutation assays amplified both DNA and RNA; for a proper evaluation of RNA templates only with these assays a DNase treatment prior cDNA synthesis was performed to remove any DNA templates. As proof-of-principle, we demonstrated that DNAse treatment successfully removed all remaining DNA content by using the PIK3CA p.H1047R mutation assay on BT20 cell line mRNA (**Figure 5**). Mutant and wild-type copies were detected in cellular mRNA specimens treated with/without DNase (+DNAse/-DNAse) after reverse transcription (+RT) (**Figures 5A** and **5C**) whereas no PIK3CA copies were generated in the sample with DNase and without RT enzyme (**Figure 5B**). Furthermore, few mutant and wild-type copies were measured when the sample was not converted into cDNA (minus RT) without DNase treatment (**Figure 5D**) indicative of limited contamination for the specimen with cellular germ-line DNA. Similarly, in EV-RNA of patient 15 which harbors a EGFR p.T790M mutation no difference was found in the number of mutant and wild-type copies between samples with and without DNase (**Figures 5E** and **F**). Finally, analysis of a genomic fragment within the intronic region of TK1 confirmed the successful removal of any residual DNA after DNase treatment in both cellular mRNA and patient EV-RNA (**Figure S5**).

Next, the PIK3CA p.H1047R mutation status analyzed by dPCR was compared between cellular mRNA and matched EV-RNA of BT20 and T47D cell lines. Mutant transcripts were present at comparable frequencies between EV-RNA and matched cellular mRNA (**Figure 6**; BT20: 30% vs 32% and T47D: 89% vs 84%, respectively). Similar results were obtained in in MBA-MB-435 for the BRAF p.V600E in cellular RNA and EV-RNA (68% vs 63%, respectively) and in MDA-MB-231 for the KRAS p.G13D mutation detected in cellular RNA and matched EV-RNA (51% vs 50%, respectively).

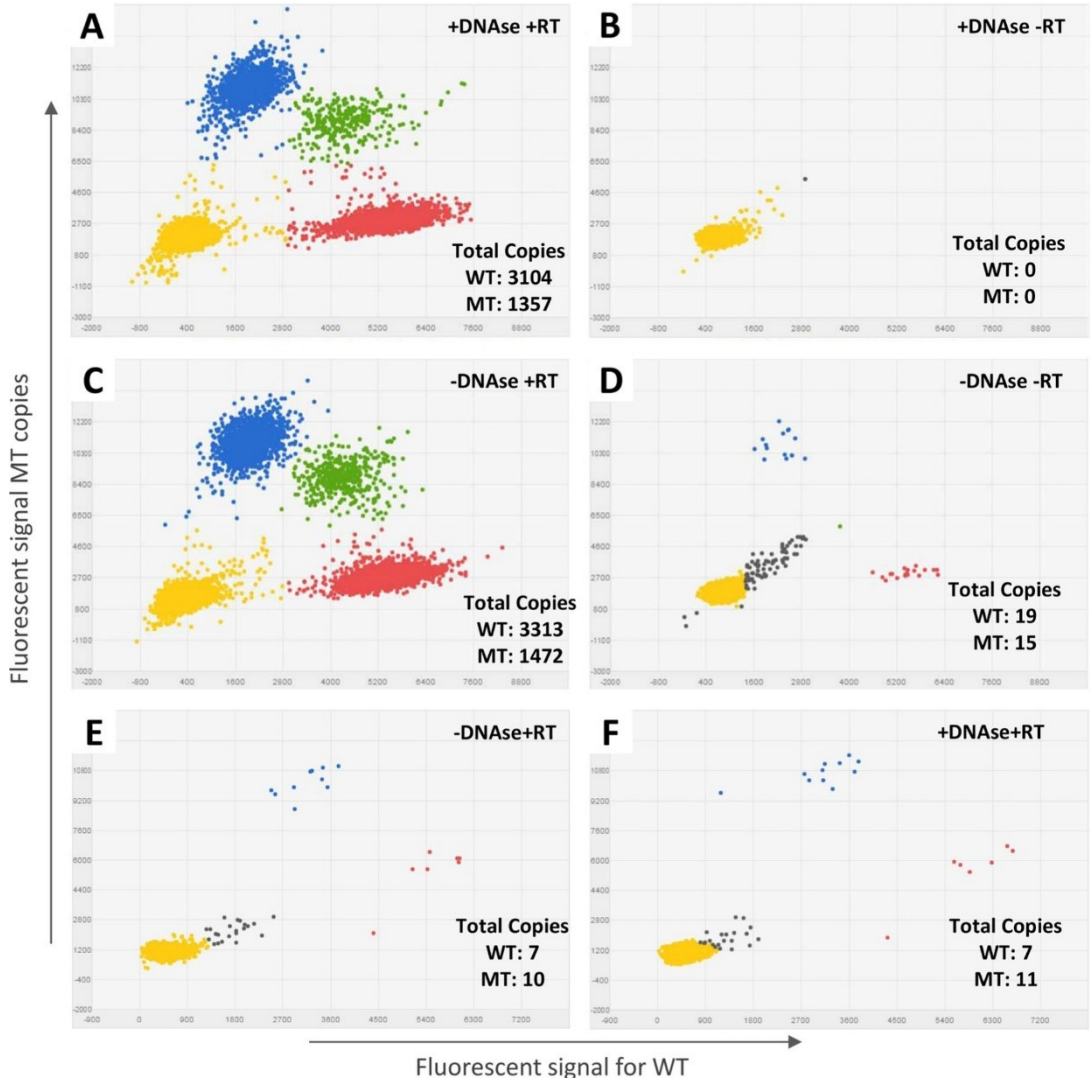


Figure 5. Mutation analysis by dPCR in (EV) RNA with and without DNAse treatment.

Dot-plots in panel a-d indicate the presence of wild-type (WT) and/ or mutant (MT) PIK3CA p.H1047R copies in BT20 mRNA treated a) with DNase and reverse transcriptase (RT), b) with DNAse and without RT, c) without DNase and with RT, d) without both DNAse and RT. Dot-plot in panel e-f indicate the presence of wild-type (WT) and/or mutant (MT) of EGFR p.T790M in patient 15 EV-RNA e) treated without DNAse and f) with DNAse treatment. Similar mutant and wild-type copies were observed in both conditions.Blue: wells with mutant copies, Red: wells with wild-type copies, Green: wells containing both wild-type and mutant copies, Yellow: empty wells, Grey: undetermined wells.

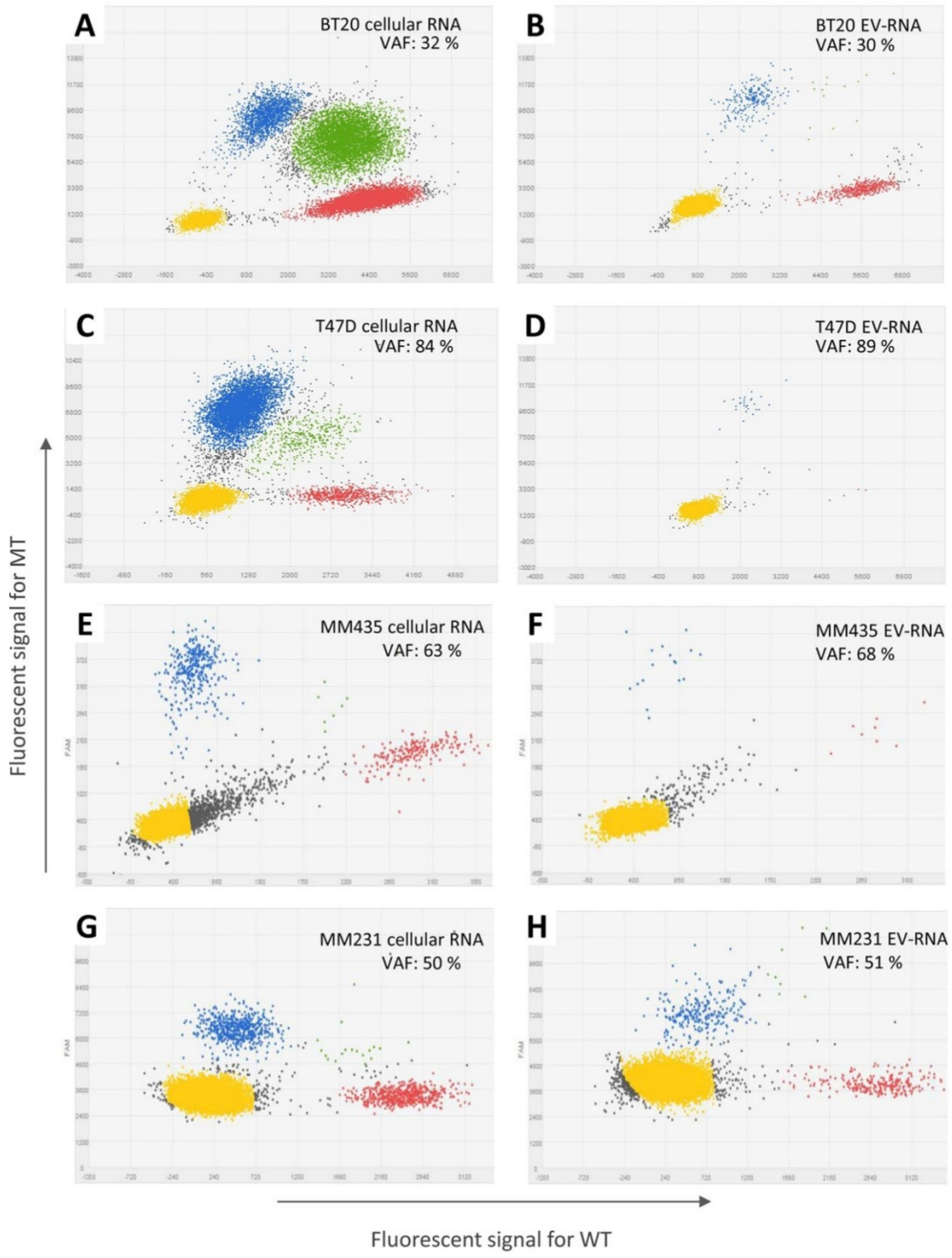


Figure 6. Mutation analysis of human breast cancer cell lines by digital PCR.

The figure shows dPCR dot plots for cellular mRNA (a, c, e and g) and matched EVRNA (b, d, f and h) indicating wild-type (WT) and mutant (MT) copies for breast cancer cell lines BT20 (PIK3CA p.H1047R), T47D (PIK3CA p.H1047R), MDA-MB-435 (BRAF p.V600E(c.1799T > A)) and MDA-MB-231 (KRAS p.G13D). Dots represent wells with mutant copies (blue), wild-type copies (red), both wild-type and mutant copies (green), empty wells (yellow), and undetermined wells (grey). VAF: Variant Allele Frequency

Analysis of gene expression and mutant transcripts in patient EV-RNA

The feasibility of above workflow to quantify gene expression together with wildtype and mutant transcript levels was also evaluated in EVs derived from minute amounts of plasma (< 1mL) from patients with various solid cancers. In all EDTA samples at 1 hour and 24 hours of the cohort I patients, expression of the reference gene GAPDH was detectable in EV-RNA with a median of 34.5×103 and 53.4×103 copies/mL plasma, respectively (**Figure 3C**). Up to ten-fold lower GAPDH levels (average median 3.06×103 copies/mL plasma) were measured in both BCT and CellSave tubes independent of their processing time-point (**Figure 3C**). In addition, no significant increase was observed in GAPDH copies/mL plasma in EDTA-blood samples processed within 24 hours (Figure S2B). For cohort II, EDTA blood from patients collected within 24 hours had a median of ~24.4×103 of GAPDH copies/mL plasma, which was 1.55-fold higher than the levels measured in plasma derived from HBDs (**Figure 3D**).

Next, the number of transcripts was established by dPCR and gene expression assay for the target gene in which a mutation was reported (**Table 1** and **Table S1**). Target gene copies were only detected in EDTA tubes (**Table 1**) but not in BCT nor in CellSave tubes (data not shown). For this reason, only results obtained in EDTA tubes of cohort I and II were reported.

Gene transcripts in EV-RNA were quantified with dPCR using both mutation and expression assay for the target gene with a somatic variant (**Table 1, Figures S6** and **S7**). Both types of assays detected at least 100 copies per mL plasma of target gene EV-RNA transcripts in all cases, except for patients 2 and 13 with no or very few copies observed. Mutant target gene transcripts ranging from 34 to 288 copies per mL plasma, on the other hand, were only found in EV-RNA from patients 5 (KRAS p.G13D),15 (EGFR p.T790M) and 21 (KRAS p.G13D) with VAFs of 15%, 24% and 2.2%, respectively (**Table 1** and **Figure 7**).

The 96 gene expression profiling in EV-RNA was retrospectively performed in EV-RNA of 6 patients (2 NSCLC, 2 CRC, 1BC, 1 cholangiocarcinoma) with enough material for the analysis and compared with those of the cancer cell lines and of leucocytes from HBDs. Hierarchical cluster analyses of these expression profiles for all genes but also for above 8 genes enriched in EVs, showed that EV-RNAs from patients clustered more closely to the HBDs than to the cancer cell lines. Only a few genes were specifically expressed in patients with (ct)EVs and in EVs of cancer cell lines. These genes included AGR2, KRT17, SPDEF, and LAD1 which were expressed in EV-RNA of patients 5 and 15 and of all cancer cell lines, but not in HBDs (**Figure 4C**).

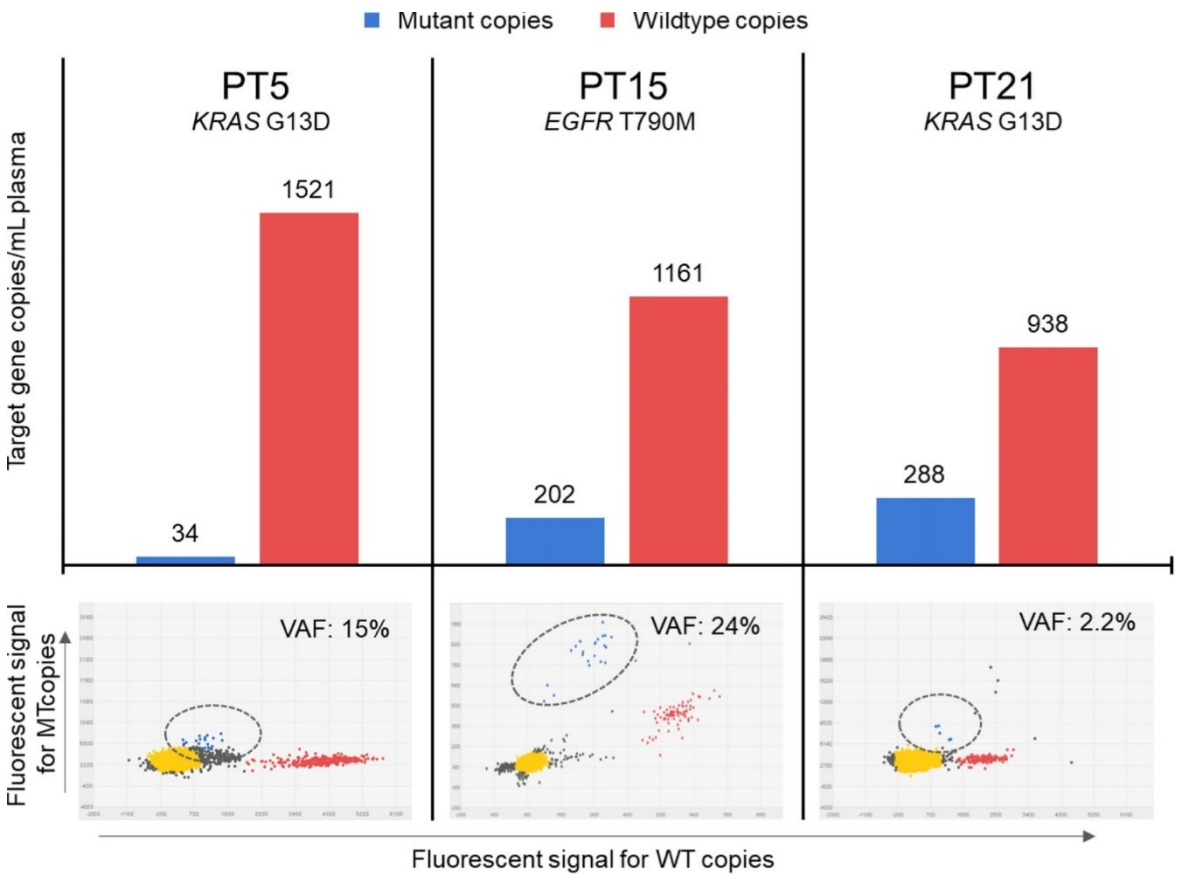


Figure 7. Patients in which mutant target gene copies were detected in EV-RNA.

The top part of the figure show blue bars which indicate the number of mutant copies/mL plasma and red bars that indicate the number of wild-type copies/mL plasma. Additionally, the bottom section of the figure show the results of the dPCR including the variant allele frequencies. Blue: represent mutant copies, Red: represent wild-type copies, Yellow: represent empty wells, Grey: represent undetermined wells, VAF: Variant Allele Frequency.

Discussion

Even though cfDNA has recently shown great promise, cfDNA analyses does not provide information on gene and protein expression which EVs likely can. For this purpose, we developed a workflow to isolate and characterize ctEVs d erived from small amounts of plasma (< 1mL) from 20 metastatic cancer patients with different types of solid tumors.

First, EV preparations of plasma from patients and HBDs were characterized to provide evidence that EVs were harvested with our workflow. It has already been reported that EVs were harvested from plasma of ovarian and prostate cancer patients using the membrane-based affinity binding step of the Exo (RNA)Easy kit and that NTA, TEM and immune assays can be used to characterize these EVs[27-28].

To have an additional quantitative measure of the EVs in the analyzed specimens, we used the CD9 TRIFic assay to measure the CD9 transmembrane protein, which is enriched up to 10 times in EVs compared to other particles [24, 29-32]. The advantage of the CD9 TRIFic assay is that it can be easily used without plasma purification and is not hampered by other non-EV particles such as lipid particles and protein complexes. Furthermore, it is not restricted by vesicle size compared to other methods for EV quantification such as flow cytometry, flow immuno-detection (LFIA), nanoparticle tracking and tunable resistive pulse sensing [24, 33, 34]. The measured CD9 levels varied considerably between conditioned cell line media (**Figure S1**), cancer patients and healthy donors but confirmed observations reported previously [24]. We also showed that processing time after blood draw affects CD9 levels in plasma samples (**Figure S2A).** Noteworthy, CD9 levels were higher in all plasma samples processed at a later time point, suggesting that EVs are released by (blood) cells [35] during storage after venipuncture but before blood was processed into plasma. Finally, higher CD9 levels were measured (although not significant) in plasma of patients with cancer compared to HBDs, which was also previously described for prostate cancer patients [36]. Our study now demonstrated similar FLOT1 levels in EV preparations from patients and HBDs. The low levels of FLOT1 may suggest that FLOT1 is not highly expressed on the surface of a vesicle but rather on the inside as FLOT1 is found on the inner membrane of a cell. Therefore, we might speculate that during the formation of the EV's by budding of the cell membrane FLOT1 remains inside an EV. Most studies performed involving measurement of FLOT1 perform this by means of western blotting, which involves a form of EV lysis making it possible to obtain higher amounts of FLOT1 [37 - 40]. Further studies are necessary to confirm these findings. Furthermore, higher APOB levels were observed between patient and

HBD EV preparations and are only slightly lower compared to APOB levels directly measured in plasma (data not shown) indicating that the used method for EV purification still co-isolated a high amount of lipoproteins. This might hamper the recovery of EV's from plasma and negatively influence detection of mutations in EV-RNA. Nonetheless, EpCAM protein levels were significantly higher in patients' EVs advocating for enrichment of ctEV using EpCAM as a target for capture. This might improve the purity of the EV preparations and also our ability to detect mutations due to a higher rate of ctEV recovery. However, further studies are needed to verify our findings and EpCAM based capturing methodologies should be explored.

Several studies reported that EVs are involved in tumorigenesis, proliferation, drug resistance, angiogenesis and the development of pre-metastatic niches [19, 41-44]. The observation of EV-derived RNA being translated into functional proteins in recipient cells shows that (tumor) cells use EVs to deliver information to other cells [45-47]. Our gene expression profiling of cellular mRNA and EV-RNA demonstrated that EVs carry multiple gene transcripts and cluster with their matched cellular counterpart. This suggests that EV-RNA express molecular (sub)type specific genes as previously shown by others [48,49]. Interestingly, 8 genes were more abundant in EV-RNA compared to cellular RNA which might be an indication of selective enrichment (e.g. several *KRTs*) (e.g. several *KRTs*) and/or exclusion of certain transcripts from EVs. Two of these EV-RNA abundant genes (*KRT17*, *SPDEF*) were also expressed in patients 5 and 15 with ctEVs whereas not expressed in healthy blood donor EV-RNA and mRNA. The gene *SPDEF* was previously reported as part of a urine exosome gene expression assay which also includes the genes *ERG* and *PCA3*. This 3-gene assay was reported to discriminate high-grade prostate cancer from low-grade and benign disease [50]. Likewise, *KRT17* and *SPDEF* might be used to discriminate between patients with high and low tumor-load as indicated by the measured primary tumor percentage and cfDNA VAF (**Table 1**). Additionally, our findings suggest that such an EV gene expression assay might also be applicable in plasma EV from patients with other malignancies. However, we did not find a real difference between EVs from patients and those derived from healthy donors while more differences were observed with EVs from tumor cell lines. We might speculate that EVs are mostly derived from healthy tissue as previously described by *Mitchell et al.* These authors showed no real difference in EVs quantity in urinary-exosome between healthy men and those with locally advanced disease[51]. However, further studies are needed to validate our findings as well as to develop methods to enrich for tumor-derived EVs.

The use of EVs as liquid biopsy biomarker for cancer patients is challenged by pre-analytical factors, like time to process blood into plasma and type of blood collection tubes used [7, 52-54].

Several studies have shown the impact of anticoagulants (such as sodium-citrate, EDTA or heparin) on EV analysis outcome [7, 54]. Our aim was the simultaneous analysis of both cfDNA and EV-RNA. Therefore, we investigated EDTA, BCT, and CellSave tubes with the latter two tubes showing stable cfDNA quality and quantity in blood processed into plasma up to 24 hours after venipuncture [23]. However, the results of our workflow show that BCT and CellSave tubes are less suited for analyses of EV-RNAs. Recently, BCT-RNA tubes were reported to stabilize cell-free RNA (cfRNA) [55], which might also be feasible for EV-RNA but due to our retrospective study, was not evaluated.

Next, we showed in EDTA tubes the impact of time delay between blood sampling and plasma processing on EVs and EV-RNA. Ideally, blood should be processed into plasma immediately after blood draw, which is not feasible in daily clinical practice and in multicenter clinical trials. We observed increased EV-RNA *GAPDH* transcripts with time to process blood into plasma after blood draw. This increase was previously also reported for cfRNA copies of *GAPDH* and beta-2-microglobulin (*B2M*) [55], showing that processing time affects downstream EV-RNA. However, the observed increase of *GAPDH* transcripts within 24 hours was moderate and not significantly different compared to the number of *GAPDH* transcripts after 1 hour processing. The 24 hours processing of blood is more feasible with current clinical practice.

Previous studies detected tumor-specific mutations and translocations in DNA and RNA from EVs, respectively [56-59]. We now evaluated EV-RNA of 20 cancer patients for somatic variants found in tissue and in 17 cases of these also in plasma cfDNA. The sensitivity for mutation detection in plasma EV-RNA (3/20 patients; 15%) was considerably lower than for cfDNA (17/20 patients; 75%). This discrepancy in mutation detection between cfDNA and EV-RNA can be partially due to differences in time of plasma storage and analyses, plasma input used (i.e. less than 1mL was available and evaluated for EV-RNA whereas 3mL plasma was used for cfDNA analyses[23]), on the presence target gene copies and whether a mutation is expressed or not. Mutation detection in EV-RNA might improve with the enrichment of cancer- and/or tissue-specific EVs, when more plasma is available for such analyses. Moreover, not all hotspot mutations in DNA result in mutant transcripts carried by EVs, which might also explain the discrepancy in detection frequency between ctDNA and ctEV. In this context, we have used a very strict definition for ctEV by indicating EVs that carry mutations in their EV-RNA. In fact, although our EpCAM ELISA analyses showed many more cases EpCAM-positive EVs, which could have been defined as ctEV, most of them did not carry detectable mutant transcripts. Interestingly, mutant transcripts were detected in all patients with high amounts of ctDNA (>20% ctDNA; patients 5, 15 and 21) except

one (patient 8). This suggests that also the tumor load in plasma might be important for successful mutant transcript detection. Further studies are needed to determine whether mutant transcripts are detectable in patients with low amounts of ctDNA, without or after enrichment of cancer -specific EVs.

Nowadays, several studies have highlighted the clinical value of EVs in providing information for real-time monitoring of disease due to their minimal invasiveness as well as the opportunity to characterize the status of the tumor by using their content, which includes proteins, DNA and RNA. In this context, mutated mRNAs in plasma EVs are currently being used for the assessment of both hematological as well as solid tumors, such as prostate, lung and other solid tumors [60]. For example, mutation in the tumor-specific mRNA of epidermal growth factor receptor was detected in EVs from serum of patients with glioblastoma [14]. EV-RNAs are therefore a snapshot of the content and state of the cells that secrete them. Compared to circulating RNA (cfRNA), RNAs cargo in extracellular vesicles are quite safe from degradation by RNases enzymes even if it is still unclear what percentage of RNA cargo in EVs is functional or nor when transferred in the recipient cell [60].

Moreover, the antigenic markers on the EVs surface make it possible to discriminate the cells from which they were derived, allowing enrichment of vesicles from a particular tissue source, such as a tumor tissue. Many new sensitive technologies, such as digital PCR (dPCR) or NGS are being used to enhance detection of specific RNA species in extracellular vesicles [61]. However, the fact that with our work we cannot really discriminate the subset of extracellular vesicles derived from a cancer tissue from other vesicles is a limit for their clinical application. Similarly, an issue concerns the choice of the best method to isolate and characterize cancer EVs from all the other vesicles which are released from the (normal) cell sources [62]. Nonetheless, our study shows a relatively fast method for obtaining EVs, generating expression profile and perform mutation analysis (with the known present limitations). However, to overcome the present limitations and enhance the clinical application of EVs for tumor management, further investigations are still needed.

Conclusions

We provide a workflow for the detection of ctEVs by a spin column-based generic isolation of EVs. The workflow was followed by PCR-based measurements of gene expression and mutant transcripts in EV-RNA derived from cancer patients' blood plasma processed within 24 hours. Tumor-specific mutations in blood, however, were less often observed in EV-RNA than in cfDNA.

References

1. Xiao, D., et al., *Identifying mRNA, microRNA and protein profiles of melanoma exosomes.* PLoS One, 2012. **7**(10): p. e46874.
2. Gallo, A., et al., *The majority of microRNAs detectable in serum and saliva is concentrated in exosomes.* PLoS One, 2012. **7**(3): p. e30679.
3. Peinado, H., et al., *Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET.* Nat Med, 2012. **18**(6): p. 883-91.
4. Lau, C., et al., *Role of pancreatic cancer-derived exosomes in salivary biomarker development.* J Biol Chem, 2013. **288**(37): p. 26888-97.
5. Choi, D.S., et al., *Proteomic analysis of microvesicles derived from human colorectal cancer ascites.* Proteomics, 2011. **11**(13): p. 2745-51.
6. Enderle, D., et al., *Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method.* PLoS One, 2015. **10**(8): p. e0136133.
7. Gyorgy, B., et al., *Improved circulating microparticle analysis in acid-citrate dextrose (ACD) anticoagulant tube.* Thromb Res, 2014. **133**(2): p. 285-92.
8. Budnik, V., C. Ruiz-Canada, and F. Wendler, *Extracellular vesicles round off communication in the nervous system.* Nat Rev Neurosci, 2016. **17**(3): p. 160-72.
9. Kahlert, C. and R. Kalluri, *Exosomes in tumor microenvironment influence cancer progression and metastasis.* J Mol Med (Berl), 2013. **91**(4): p. 431-7.
10. Heijnen, H.F., et al., *Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules.* Blood, 1999. **94**(11): p. 3791-9.
11. Lobb, R.J., et al., *Optimized exosome isolation protocol for cell culture supernatant and human plasma.* J Extracell Vesicles, 2015. **4**: p. 27031.
12. Thery, C., M. Ostrowski, and E. Segura, *Membrane vesicles as conveyors of immune responses.* Nat Rev Immunol, 2009. **9**(8): p. 581-93.
13. Azmi, A.S., B. Bao, and F.H. Sarkar, *Exosomes in cancer development, metastasis, and drug resistance: a comprehensive review.* Cancer Metastasis Rev, 2013. **32**(3-4): p. 623-42.
14. Skog, J., et al., *Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers.* Nat Cell Biol, 2008. **10**(12): p. 1470-6.
15. Thery, C., L. Zitvogel, and S. Amigorena, *Exosomes: composition, biogenesis and function.* Nat Rev Immunol, 2002. **2**(8): p. 569-79.

16. Miranda, K.C., et al., *Nucleic acids within urinary exosomes/microvesicles are potential biomarkers for renal disease.* Kidney Int, 2010. **78**(2): p. 191-9.
17. Taylor, D.D. and C. Gercel-Taylor, *MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer.* Gynecol Oncol, 2008. **110**(1): p. 13-21.
18. Nolte-'t Hoen, E.N., et al., *Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions.* Nucleic Acids Res, 2012. **40**(18): p. 9272-85.
19. Valadi, H., et al., *Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells.* Nat Cell Biol, 2007. **9**(6): p. 654-9.
20. Smirnov, D.A., et al., *Global gene expression profiling of circulating tumor cells.* Cancer Res, 2005. **65**(12): p. 4993-7.
21. Schageman, J., et al., *The complete exosome workflow solution: from isolation to characterization of RNA cargo.* Biomed Res Int, 2013. **2013**: p. 253957.
22. Huang, X., et al., *Characterization of human plasma-derived exosomal RNAs by deep sequencing.* BMC Genomics, 2013. **14**: p. 319.
23. van Dessel, L.F., et al., *Application of circulating tumor DNA in prospective clinical oncology trials: standardization of pre-analytical conditions.* Mol Oncol, 2017.
24. Thery, C., et al., *Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines.* J Extracell Vesicles, 2018. **7**(1): p. 1535750.
25. Duijvesz, D., et al., *Immuno-based detection of extracellular vesicles in urine as diagnostic marker for prostate cancer.* Int J Cancer, 2015. **137**(12): p. 2869-78.
26. Sieuwerts, A.M., et al., *Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR.* Breast Cancer Res Treat, 2009. **118**(3): p. 455-68.
27. Sieuwerts, A.M., et al., *mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients.* Clin Cancer Res, 2011. **17**(11): p. 3600-18.
28. Zhang W, et al., P*roteomics profiling of plasma exosomes in epithelial ovarian cancer: a potential role in the coagulation cascade, diagnosis and prognosis.* Int J Oncol. 2019;**54**(5):p. 1719–33.
29. Worst TS, et al. *miR-10a-5p and miR-29b-3p as Extracellular Vesicle-Associated Prostate Cancer Detection Markers.* Cancers (Basel). 2019;**12**(1).

30. Thery C, et al., *Molecular characterization of dendritic cell-derived exosomes. Selective accumulation of the heat shock protein hsc73.* J Cell Biol. 1999;**147**(3):p. 599–610.

31. Escola JM, et al., *Selective enrichment of tetraspan proteins on the internal vesicles of multivesicular endosomes and on exosomes secreted by human B- lymphocytes.* J Biol Chem. 1998;**273**(32):p. 20121–7.

32. Denys H, et al., T*he isolation of morphologically intact and biologically active extracellular vesicles from the secretome of cancer-associated adipose tissue.* Cell Adh Migr. 2017;**11**(2):p. 196–204

33. Andreu Z, et al., *Tetraspanins in extracellular vesicle formation and function. Front Immunol.* 2014;**5**:442.

34. Momen-Heravi F, et al., *Alternative methods for characterization of extracellular vesicles. Front Physiol.* 2012;**3**:354.

35. Oliveira-Rodriguez M, et al., *Development of a rapid lateral flow immunoassay test for detection of exosomes previously enriched from cell culture medium and body fluids*. J Extracell Vesicles. 2016;**5**:31803.

36. Tarrant JM, et al., *Tetraspanins: molecular organisers of the leukocyte surface. Trends Immuno*l. 2003;**24**(11):610–7.

37. Soekmadji C, et al. *Modulation of paracrine signaling by CD9 positive small extracellular vesicles mediates cellular growth of androgen deprived prostate cancer*. Oncotarget. 2017;**8**(32):52237–55.

38. Sjoqvist S, et al*., Exosomes derived from clinical-grade oral mucosal epithelial cell sheets promote wound healing.* J Extracell Vesicles. 2019;8(1):1565264.

39. Kowal J, et al., *Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes*. Proc Natl Acad Sci U S A. 2016;**113(**8):E968–77.

40. Koh YQ, et al. *Proteome profiling of exosomes derived from plasma of heifers with divergent genetic merit for fertility.* J Dairy Sci. 2018;**101**(7):6462–73.

41. Hong, B.S., et al., *Colorectal cancer cell-derived microvesicles are enriched in cell cycle-related mRNAs that promote proliferation of endothelial cells.* BMC Genomics, 2009. **10**: p. 556.

42. Challagundla, K.B., et al., *Exosome-mediated transfer of microRNAs within the tumor microenvironment and neuroblastoma resistance to chemotherapy.* J Natl Cancer Inst, 2015. **107**(7).

43. Kosaka, N., et al., *Neutral sphingomyelinase 2 (nSMase2)-dependent exosomal transfer of angiogenic microRNAs regulate cancer cell metastasis.* J Biol Chem, 2013. **288**(15): p. 10849-59.

44. Fong, M.Y., et al., *Breast-cancer-secreted miR-122 reprograms glucose metabolism in premetastatic niche to promote metastasis.* Nat Cell Biol, 2015. **17**(2): p. 183-94.

45. Pegtel, D.M., et al., *Functional delivery of viral miRNAs via exosomes.* Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6328-33.

46. Zhang, Y., et al., *Secreted monocytic miR-150 enhances targeted endothelial cell migration.* Mol Cell, 2010. **39**(1): p. 133-44.

47. Kosaka, N., et al., *Secretory mechanisms and intercellular transfer of microRNAs in living cells.* J Biol Chem, 2010. **285**(23): p. 17442-52.

48. Jenjaroenpun, P., et al., *Characterization of RNA in exosomes secreted by human breast cancer cell lines using next-generation sequencing.* PeerJ, 2013. **1**: p. e201.

49. Lazaro-Ibanez, E., et al., *Distinct prostate cancer-related mRNA cargo in extracellular vesicle subsets from prostate cell lines.* BMC Cancer, 2017. **17**(1): p. 92.

50. McKiernan, J., et al., *A Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer at Initial Biopsy.* JAMA Oncol, 2016. **2**(7): p. 882-9.

51. Mitchell, P.J., et al., *Can urinary exosomes act as treatment response markers in prostate cancer?* J Transl Med, 2009. **7**: p. 4.

52. Baek, R., et al., *The impact of various preanalytical treatments on the phenotype of small extracellular vesicles in blood analyzed by protein microarray.* J Immunol Methods, 2016. **438**: p. 11-20.

53 Lacroix, R., et al., *Standardization of pre-analytical variables in plasma microparticle determination: results of the International Society on Thrombosis and Haemostasis SSC Collaborative workshop.* J Thromb Haemost, 2013.

54. Wisgrill, L., et al., *Peripheral blood microvesicles secretion is influenced by storage time, temperature, and anticoagulants.* Cytometry A, 2016. **89**(7): p. 663-72.

55. Fernando, M.R., et al., *Stabilization of cell-free RNA in blood samples using a new collection device.* Clin Biochem, 2012. **45**(16-17): p. 1497-502.

56. Fabienne Kunz, K.R., Nils Von Neuhoff, Karl Welte, Dirk Reinhardt and Basant Kumar Thakur, *Evaluation of Genetic Mutations in Plasma Exosomes As a Biomarker Tool for Detecting the Clonal Evolution of Leukemic Cells in Pediatric AML.* Blood, 2016. **128**(22): p. 1711.

57. Kahlert, C., et al., *Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer.* J Biol Chem, 2014. **289**(7): p. 3869-75.

58. San Lucas, F.A., et al., *Minimally invasive genomic and transcriptomic profiling of visceral cancers by next-generation sequencing of circulating exosomes.* Ann Oncol, 2016. **27**(4): p. 635-41.

59. Takahashi, K., et al., *Analysis of extracellular RNA by digital PCR.* Front Oncol, 2014. **4**: p. 129.

60. O'Brien, K., et al., *RNA delivery by extracellular vesicles in mammalian cells and its applications.* Nat Rev Mol Cell Biol, 2020. **21**(10): p. 585-606.

61. Shao, H., et al., *New Technologies for Analysis of Extracellular Vesicles.* Chem Rev, 2018. **118**(4): p. 1917-1950.

62. Zhu L, et al. *Isolation and characterization of exosomes for cancer research. J Hematol Oncol.* 2020;**13**(1):152.

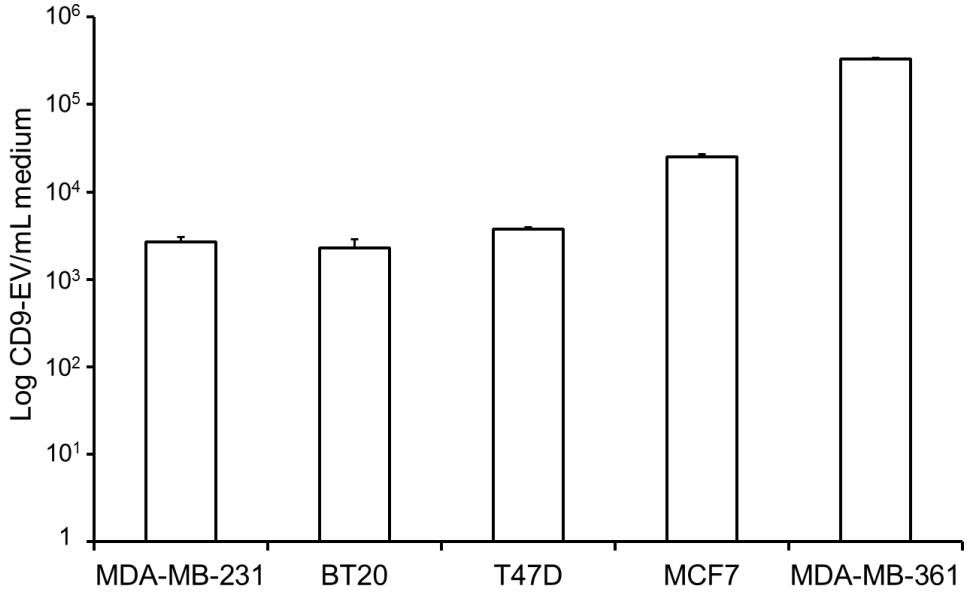


Figure S1. Analyses of CD9-EV levels by TRIFic™ exosome assay.

The bar chart shows the log of CD9-EV counts per mL cell culture medium measured in five breast cancer cell lines.

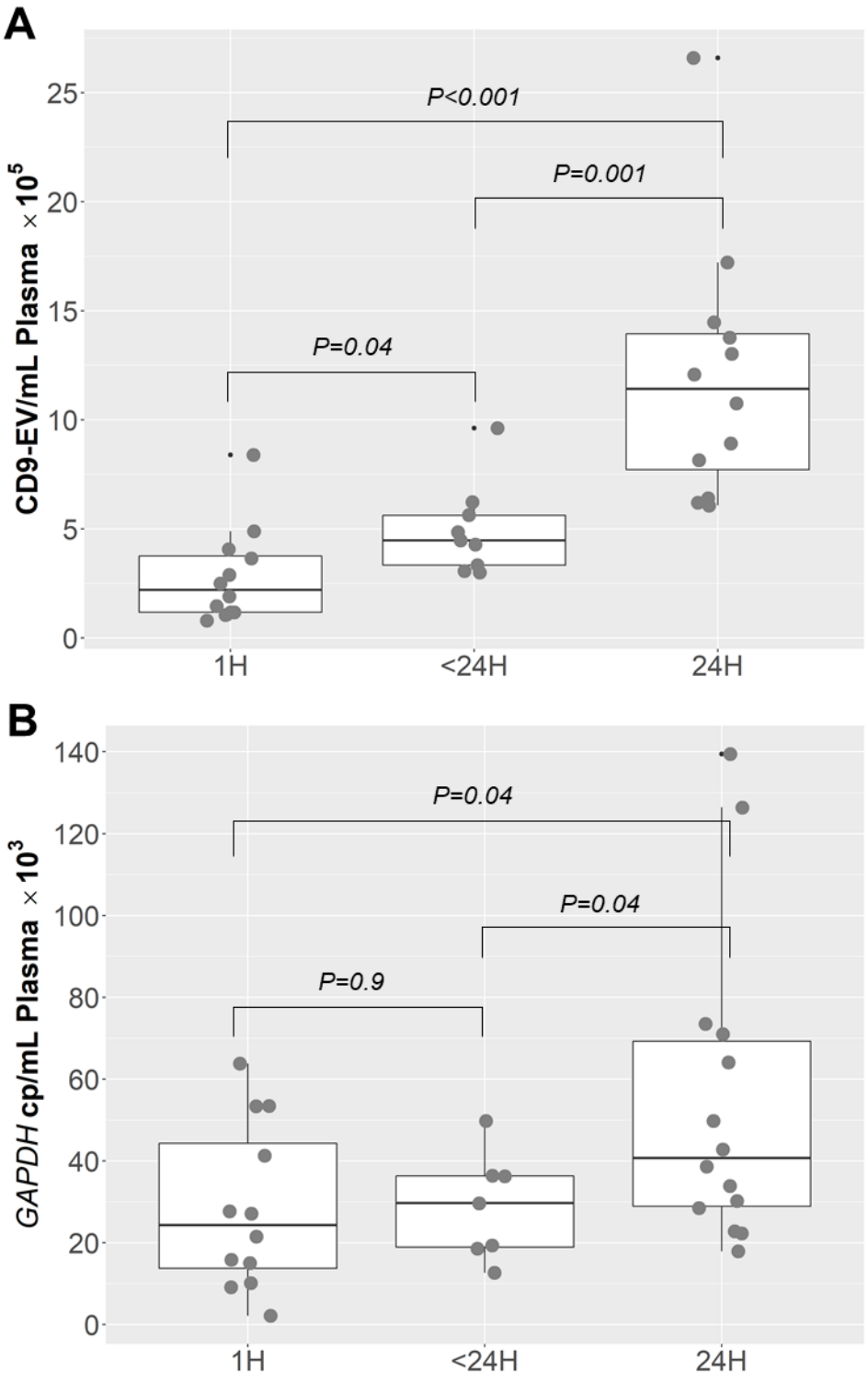


Figure S2 - TRiFIC™ Analysis of CD9-EV in plasma of cancer patients and *GAPDH* copies in their EV-RNA.

The boxplots shows in A) CD9-EV levels per mL plasma and in B) the *GAPDH* copies/mL plasma measured in EDTA tubes and processed at three time points (1 hour, within 24 hour and 24 hours) for 8 patients from each cohort 1 (at 1h and 24h) and cohort 2 (<24h). All measurements for CD9-EV were performed in duplicate.

BT-20 **(ATCC® HTB-19)**	Amelogenin: X CSF1PO: 12 D13S317: 11 D16S539: 11,14 D5S818: 12 D7S820: 10 THO1: 7,9.3 TPOX: 11 vWA: 16,17
T-47D ***(ATCC® HTB-133)***	Amelogenin: X CSF1PO: 11,13 D13S317: 12 D16S539: 10 D5S818: 12 D7S820: 11 THO1: 6 TPOX: 11 vWA: 14
MCF7 ***(ATCC® HTB-22)***	Amelogenin: X CSF1PO: 10 D13S317: 11 D16S539: 11,12 D5S818: 11,12 D7S820: 8,9 THO1: 6 TPOX: 9,12 vWA: 14,15
MDA MB 231 ***(ATCC® HTB-26)***	Amelogenin: X CSF1PO: 12,13 D13S317: 13 D16S539: 12 D5S818: 12 D7S820: 8,9 THO1: 7,9.3 TPOX: 8,9 vWA: 15,18
MDA MB 435 ***(ATCC® HTB-131)***	Amelogenin: X CSF1PO: 11 D13S317: 12 D16S539: 13 D5S818: 12 D7S820: 8,10 THO1: 6,7 TPOX: 8,11 vWA: 16,18
MDA MB 361 ***(ATCC® HTB-27)***	Amelogenin: X CSF1PO: 12 D13S317: 11 D16S539: 11,12 D5S818: 10,11 D7S820: 9,12 THO1: 9.3 TPOX: 8,11 vWA: 17

Figure S3A.
Cell line STR Authentication measured by Powerplex 16 (Promega, cat: DC6531) of cell lines obtained from ATCC.

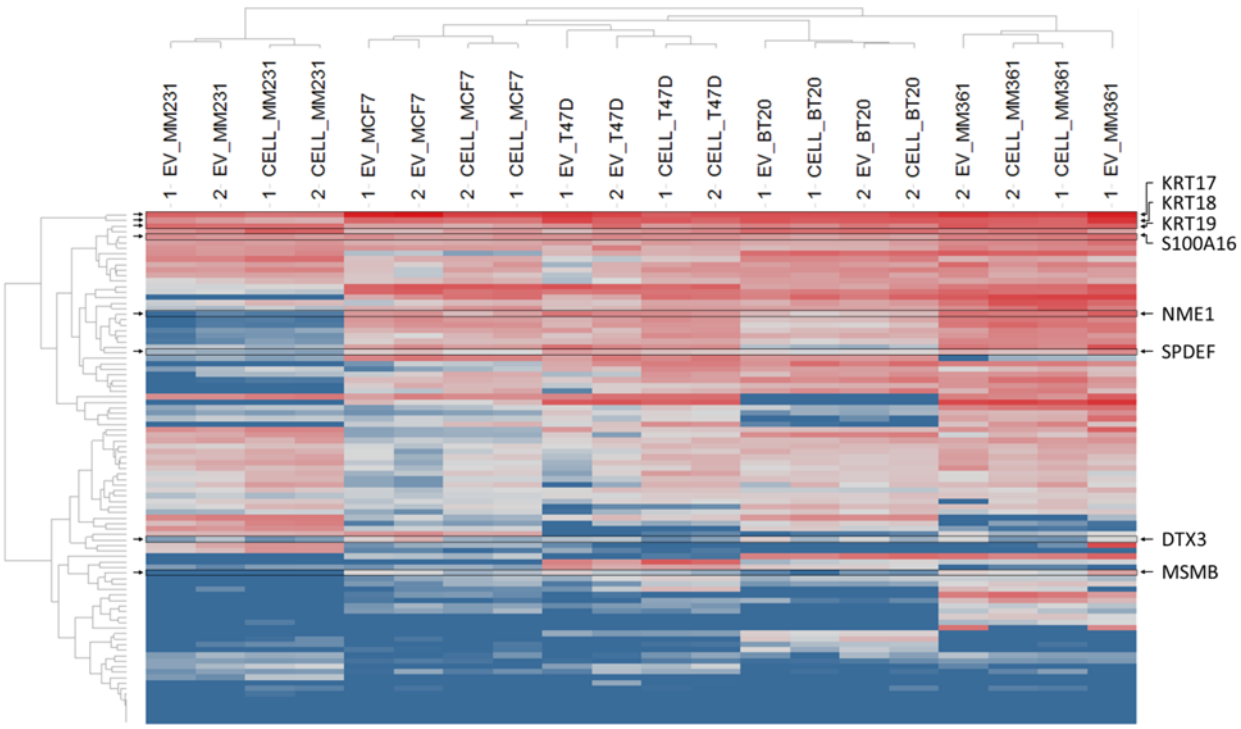


Figure S3B. Expression profiling of breast cancer cell line mRNA and EV-RNA.

TaqMan Gene Expression Assays for 96 genes were used to evaluate the RNA expression levels by real time RT-PCR. The measurements were performed in duplicate on RNA from both cell line and EVs. Complete linkage cluster analysis was performed for both cell line mRNA and matched EV-RNA and demonstrated that EV expression profiles are similar to their parental cell line profiles. Each horizontal row represents a gene, and each vertical column corresponds to a sample with numbers indicating a separate analysis (1 and 2). Arrows at the right of the figure indicate the genes which are upregulated in EVs. Expression levels are colored at median (white), above median (red) or below median (blue).

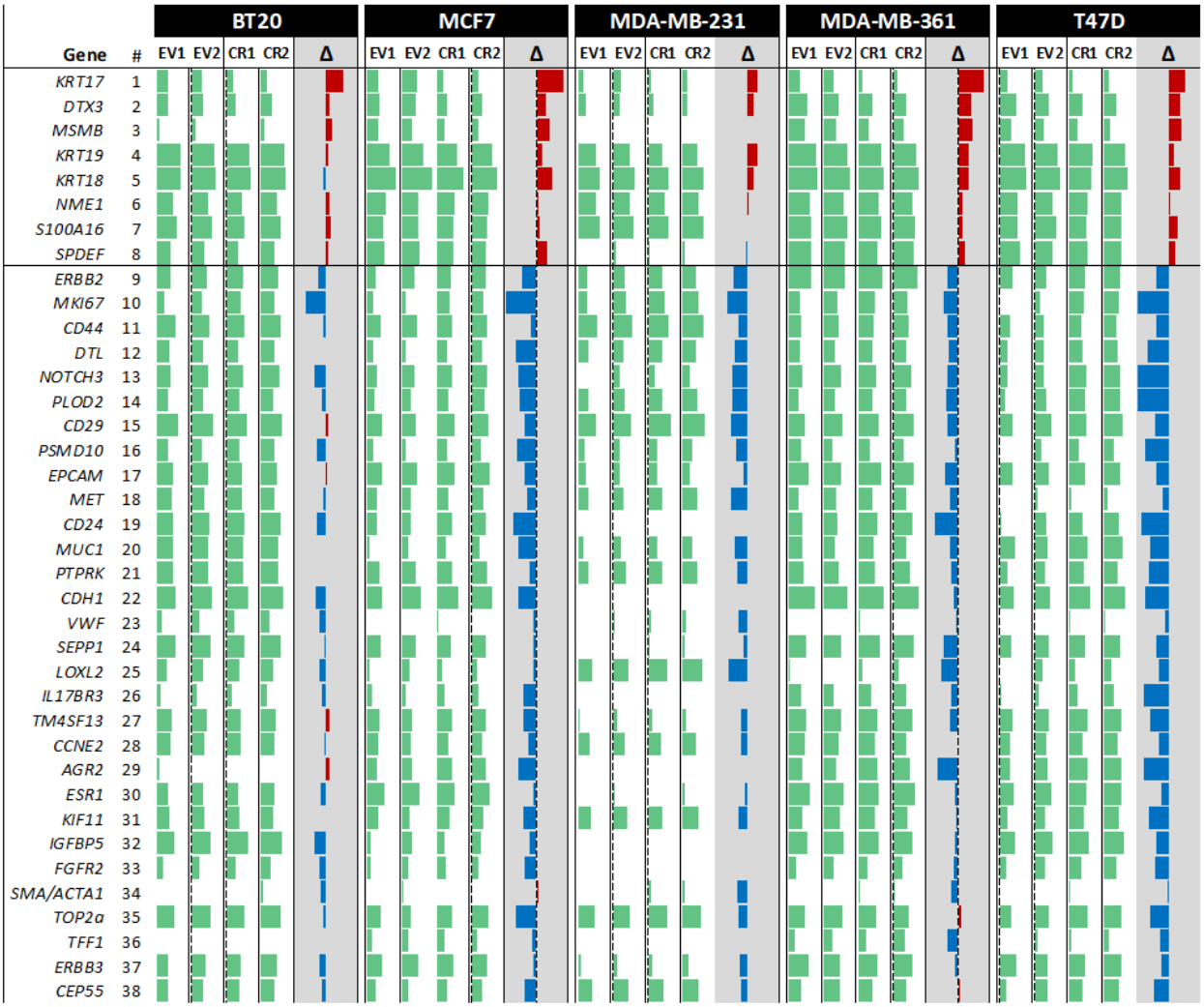


Figure S4. Differentially expressed genes in EV-RNA compared to their matched cellular mRNA.

The data are based on the independent duplicate analysis of 96 genes, but the figure shows only the 38 genes which overall were significantly different between EV-RNA and cellular RNA. Genes are indicated in the first column. The green bars for the individual cell lines represent expression levels of each gene relative to 3 reference genes (*HMBS*, *HPRT1* and *GUSB*) for two independent measurements of EV-RNA (EV1, EV2) and of cellular RNA (CR1, CR2). The difference in expression levels between EV-RNA and cellular RNA are shown by bars in red (EV>CR) or in blue (EV<CR). The first eight genes are more abundant in EV-RNA than in cellular RNA for almost all five cell lines.

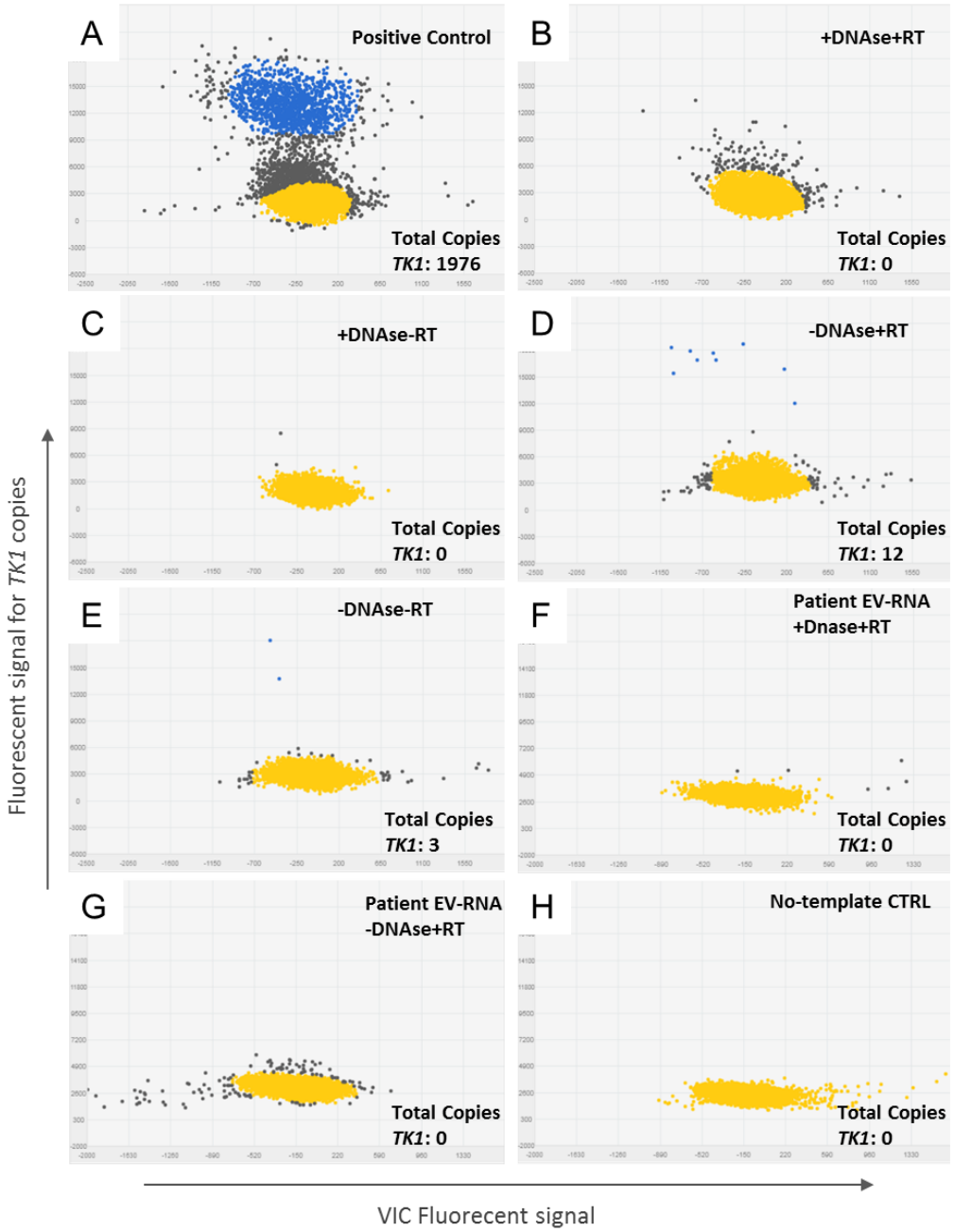


Figure S5. Analysis of an intronic region of *TK1* on cellular EV-RNA and patient EV-RNA.

The presence of DNA was evaluated in A) Celline DNA (positive control) and cellular EV-RNA treated in B) with DNAse and Reverse Transcriptase (RT); in C) with DNAse and without RT; in D) without DNAse and with RT; in E) without both DNAse and RT; and in patient EV-RNA treated in F) with DNAse and RT; in G) without DNAse and with RT; and H) show results for the no template control. No DNA was detected in both cellular and patient EV-RNA treated with DNAse. Blue: represent *TK1* copies, Yellow: represent empty wells, Grey: represent undetermined wells, VAF: Variant Allele Frequency.

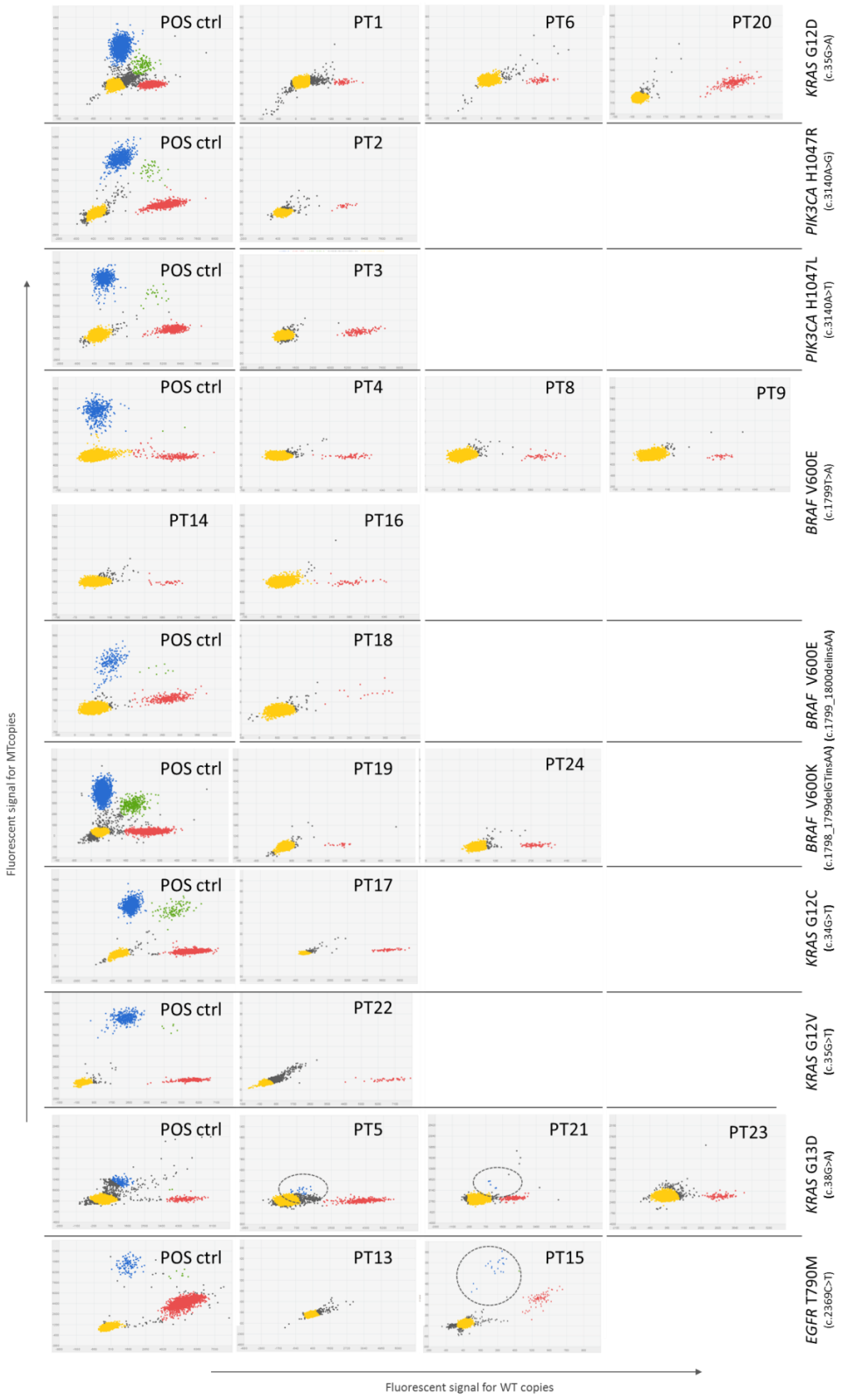


Figure S6. Digital PCR dot-plots of target genes in 20 patients.

Digital PCR dotplots of target genes in 20 patients. The dPCR-plots for each of mutation assays are presented for cohort 1 of EDTA plasma at 1 hour (12 patients: # 1-16) and for cohort 2 of EDTA plasma within 24 hours (8 patients: #17-24). Dots represent wells with mutant copies (blue), wild-type copies (red), both wild-type and mutant copies (green), empty wells (yellow), and undetermined wells (grey). PT: Patient; POS ctrl: Positive Control.

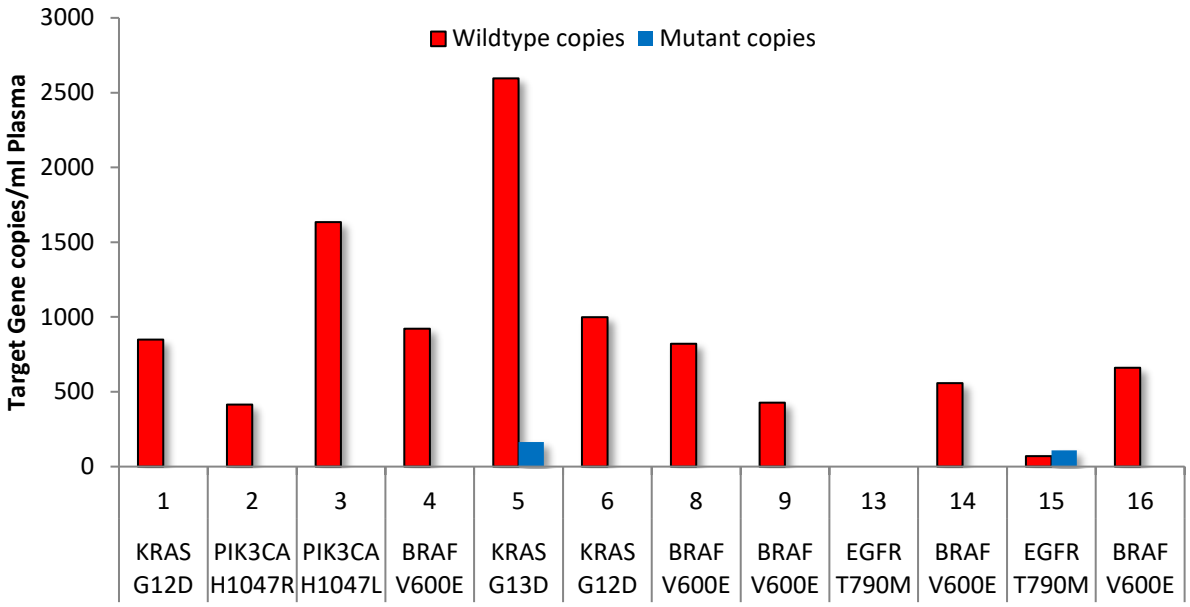


Figure S7. Mutation analyses of target genes in EV-RNA of 12 patients derived from EDTA 24 hours plasma.

Wild-type and mutant copies/mL plasma measured by dPCR from EDTA plasma processed at 24 hours after blood draw from the cohort 1 cancer patients.

Table S1. Assays used for expression analyses and for wild-type and mutant copies of target genes

Gene	Type	Nucleotide change	Amino acid change	Assay ID	Manufacturer
PIK3CA	Mutation	c.3140A>G	p.H1047R	AHPAVCD	Thermo Fisher
PIK3CA	Mutation	c.3140A>T	p.H1047L	AHLJ0TP	Thermo Fisher
KRAS	Mutation	c.35G>A	p.G12D	dHsaCP2500596	Bio-Rad
KRAS	Mutation	-	Wildtype	dHsaCP2500597	Bio-Rad
KRAS	Mutation	Screening Multiplex Assay	p.G12/G13	1863506	Bio-Rad
KRAS	Mutation	c.34G>T	p.G12C	AN9HJKW	Thermo Fisher
KRAS	Mutation	c.35G>T	p.G12V	ANAAAYM	Thermo Fisher
BRAF	Mutation	c.1798_1799delGTinsAA	p.V600K	ANNKR4W	Thermo Fisher
BRAF	Mutation	c.1799_1800delinsAA	p.V600E	AN47WF2	Thermo Fisher
BRAF (exon spanning)	Mutation	c.1799T>A	p.V600E	ANCE3VC	Thermo Fisher
BRAF	Mutation	c.1799T>A	p.V600E	dHsaCP2000027	Bio-Rad
BRAF	Mutation	-	Wildtype	dHsaCP2000028	Bio-Rad
EGFR	Mutation	c.2369C>T	p.T790M	AHRSROS	Thermo Fisher
PIK3CA	Expression	-	-	Hs00907957_m1	Thermo Fisher
KRAS	Expression	-	-	Hs00364284_g1	Thermo Fisher
BRAF	Expression	-	-	Hs00269944_m1	Thermo Fisher
EGFR	Expression	-	-	Hs00193306_m1	Thermo Fisher
GAPDH	Expression	-	-	4326317E	Thermo Fisher

Table S2. Pearson correlation coefficient R based on expression levels of a 96 gene-panel

	MM231 CELL1	MM231 EV1	MM361 CELL1	MM361 EV1	T47D CELL1	T47D EV1	BT20 CELL1	BT20 EV1	MCF7 CELL1	MCF7 EV1
MM231 CELL2	0.996	0.964	0.245	0.295	0.309	0.295	0.485	0.464	0.422	0.409
MM231 EV2	0.977	0.980	0.288	0.358	0.357	0.362	0.503	0.493	0.476	0.487
MM361 CELL2	0.229	0.259	0.991	0.856	0.768	0.655	0.587	0.583	0.776	0.699
MM361 EV2	0.246	0.301	0.905	0.881	0.684	0.618	0.516	0.523	0.718	0.689
T47D CELL2	0.347	0.370	0.763	0.667	0.993	0.855	0.662	0.641	0.800	0.721
T47D EV2	0.340	0.405	0.703	0.673	0.930	0.911	0.619	0.618	0.782	0.781
BT20C ELL2	0.442	0.455	0.607	0.512	0.650	0.545	0.988	0.970	0.669	0.621
BT20E V2	0.464	0.485	0.585	0.506	0.638	0.553	0.980	0.981	0.657	0.623
MCF7 CELL2	0.440	0.467	0.773	0.727	0.788	0.695	0.689	0.674	0.990	0.941
MCF7 EV2	0.362	0.429	0.674	0.707	0.702	0.758	0.595	0.607	0.912	0.953

1: first experiment,
2: second experiment

Table S3. Differentially expressed genes in EV-RNA compared to cell line mRNA

#	Expression[a]	Gene	P-value	Per mutation P-Value
		Upregulated genes in EV-RNA		
1	8.33	*KRT17*	2.98E-05	0.0020
2	2.56	*DTX3*	0.0006	0.0020
3	2.78	*MSMB*	0.0047	0.0078
4	2.04	*KRT19*	0.0065	0.0039
5	2.38	*KRT18*	0.0073	0.0078
6	1.28	*NME1*	0.0074	0.0078
7	1.56	*S100A16*	0.0173	0.0117
8	1.75	*SPDEF*	0.0375	0.0410
		Downregulated genes in EV-RNA		
9	0.27	*ERBB2*	9.45E-05	0.0020
10	0.06	*MKI67*	0.0004	0.0020
11	0.44	*CD44*	0.0010	0.0059
12	0.25	*DTL*	0.0011	0.0039
13	0.14	*NOTCH3*	0.0015	0.0020
14	0.18	*PLOD2*	0.0021	0.0020
15	0.33	*CD29*	0.0022	0.0078
16	0.24	*PSMD10*	0.0030	0.0039
17	0.43	*EPCAM*	0.0038	0.0117
18	0.40	*MET*	0.0039	0.0039
19	0.16	*CD24*	0.0042	0.0078
20	0.29	*MUC1*	0.0042	0.0020
21	0.38	*PTPRK*	0.0049	0.0039
22	0.30	*CDH1*	0.0073	0.0156
23	0.61	*VWF*	0.0092	0.0078
24	0.47	*SEPP1*	0.0093	0.0078
25	0.30	*LOXL2*	0.0096	0.0156
26	0.35	*IL17BR3*	0.0112	0.0078
27	0.39	*TM4SF13*	0.0118	0.0156
28	0.59	*CCNE2*	0.0121	0.0156
29	0.27	*AGR2*	0.0160	0.0313
30	0.65	*ESR1*	0.0168	0.0098
31	0.38	*KIF11*	0.0211	0.0195
32	0.49	*IGFBP5*	0.0230	0.0313
33	0.47	*FGFR2*	0.0230	0.0156
34	0.65	*SMA/ACTA1*	0.0307	0.0469
35	0.35	*TOP2α*	0.0307	0.0273
36	0.61	*TFF1*	0.0319	0.0625
37	0.55	*ERBB3*	0.0353	0.0430
38	0.46	*CEP55*	0.0402	0.0371

[a] Gene expression in EV-RNA compared to matched cell line NA

Chapter 6

An Optimized Workflow to Evaluate Estrogen Receptor Gene Mutations in Small Amounts of Cell-Free DNA

Vitale SR*, Sieuwerts AM*, Beije N, Kraan J, Angus L, Mostert B, Reijm EA, Van NM, van Marion R, Dirix LY, Hamberg P, de Jongh FE, Jager A, Foekens JA, Vigneri P, Sleijfer S, Jansen MPHM, Martens JWM

J Mol Diagn. 2019 Jan;21(1):123-137
doi: 10.1016/j.jmoldx.2018.08.010

Abstract

The detection of mutated genes in cell-free DNA (cfDNA) in plasma has emerged as an important minimally invasive way to obtain detailed information regarding tumor biology. Reliable determination of circulating tumor-derived DNA, often present at a low quantity amidst an excess of normal DNA in plasma, would be of added value for screening and monitoring of cancer patients and for hypothesis generating studies in valuable retrospective cohorts.

Our aim was to establish a workflow to simultaneously assess four hotspot estrogen receptor mutations (m*ESR1*) in cfDNA isolated from only 200 µL of plasma by means of uniplex or multiplex pre-amplification combined with digital PCR. This workflow was then applied in metastatic breast cancer (MBC) patients receiving systemic therapies for MBC. In accordance with previous studies, estrogen receptor mutations were more frequently detected in endocrine-treated MBC patients at progressive disease [34.1% (15/44)] than before the start of endocrine therapy [3.9% (2/51); *P*= 0.001]. For a subset of samples, results were compared with analysis of these mutations by Oncomine-targeted next-generation sequencing, which, although requiring a higher cfDNA input, yielded concordant results.

The data establish development and validation of a digital PCR workflow for the simultaneous detection of several tumor-derived mutations in minute amounts of cfDNA and show the potential of this workflow for use on archived volume-limited blood samples.

Introduction

Breast cancer (BC) is the most common cancer among women. Approximately 70% of BC cases are positive for the estrogen receptor (ER) [1 -3]. For patients with ER-positive metastatic BC (MBC), endocrine therapy is the preferred treatment modality. Unfortunately, approximately 40% of ER-positive MBC patients encounter swift progression after initiation of endocrine therapy and eventually virtually all MBC patients acquire resistance during treatment [1]. Mutations in the gene coding for ER (*ESR1*) have been linked to endocrine resistance and are observed in 14% to 39% of MBC patients [4-8]. Although these mutations rarely occur in primary tumors, [2, 9] they accumulate frequently during treatment [2, 10, 11], usually in patients who received prior treatment for MBC with aromatase inhibitors [6, 12, 13].

The current consensus describes a hotspot region within the ligand-binding domain of ESR1, which, if mutated, results in an amino acid substitution at positions 537 and 538 in helix 12 of the ER, closing the pocket that otherwise captures estrogen and resulting in a constitutively active ER [6, 14]. Depending on the type of endocrine therapy administered, this can result in endocrine resistance, which may be overcome by switching to agents such as fulvestrant [2, 10]. The detection of mutated *ESR1* (m*ESR1*) in MBC may play an important role in future clinical treatment decision making. However, because of the heterogeneous nature of metastases [15, 16] and because repeated tissue biopsies to observe patients in time are costly, a burden to the patient, and often difficult, if not impossible, to perform, alternative ways to assess the genetic status of MBC, including the use of liquid biopsies, are being investigated [3]. Easily accessible body fluids, such as blood plasma and serum, contain cellfree DNA (cfDNA), in which circulating tumor DNA can be detected. Because cfDNA is also released by normal cells during physiological processes (eg, apoptosis and necrosis), the discrimination between scant circulating tumor DNA derived from tumor cells among that from normal cells can be challenging [17].

Currently, research is focused on optimizing molecular methods to investigate and monitor the mutational status measured in cfDNA, and digital PCR (dPCR) has shown to be a highly sensitive approach, yielding encouraging results for the detection of m*ESR1* in MBC [3, 5, 6, 8, 11, 12, 18-25]. However, most of these studies do not take the-sometimes limited-amount of available starting material into account and typically start with at least 1 mL of plasma. To exploit plasma available from valuable retrospective studies, how to make optimal use of limited available amounts of plasma and, hence, circulating tumor DNA needs to be addressed.

Our aim was to optimize a reliable and reproducible workflow to investigate the m*ESR1* status for four hotspot mutations in minute amounts of cfDNA by dPCR and to improve cfDNA analyses from small volumes of plasma of MBC patients in general. To validate the workflow, dPCR data were compared with data obtained with a targeted nextgeneration sequencing (NGS) approach. Finally, to show the feasibility of using this workflow in a clinical setting with limited amounts of plasma available, our optimized workflow was used to evaluate the correlation between the presence of m*ESR1* and prior treatment within a retrospective cohort of MBC patients.

Materials and Methods

Study Design

Figure 1 shows a detailed overview of the study design.

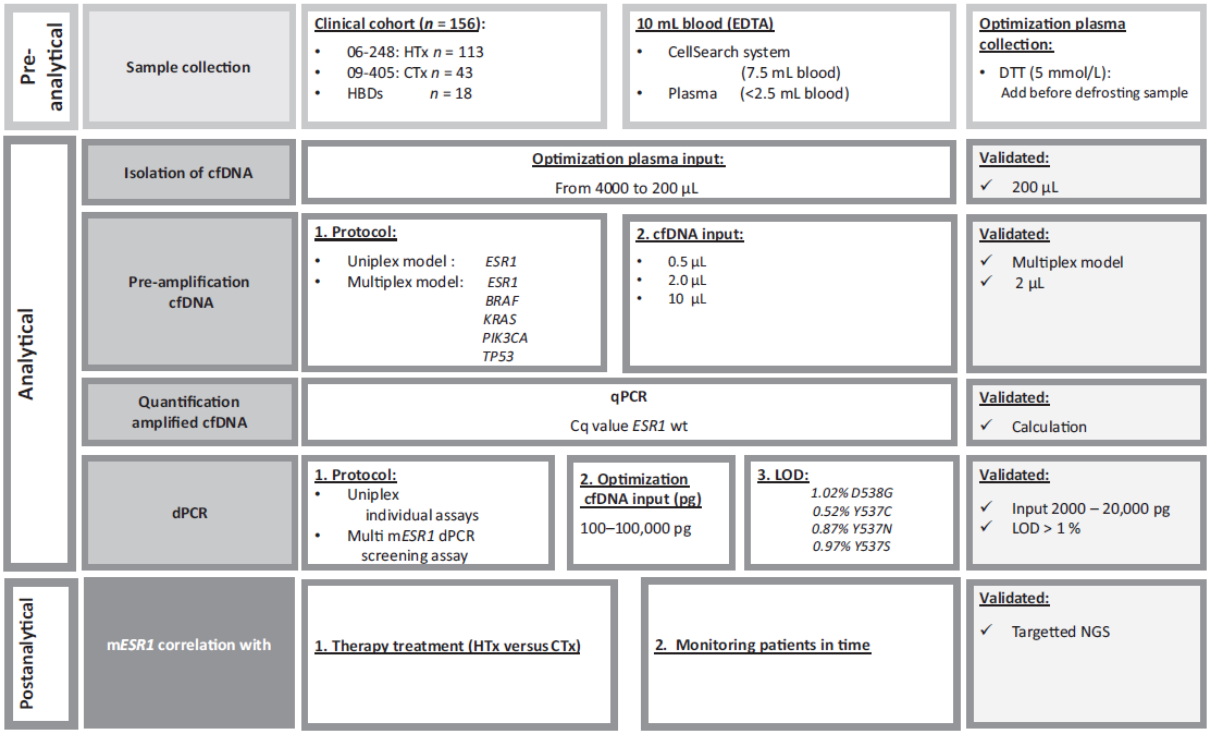


Figure 1. Workflow processing samples.

Cq, quantification cycle; cfDNA, cell-free DNA; CTx, chemotherapy; dPCR, digital PCR; DTT, dithiothreitol; HBD, healthy blood donor; HTx, endocrine therapy; LOD, limit of detection; NGS, next-generation sequencing; qPCR, quantitative PCR.

Study Cohorts

A retrospective study was performed in MBC patients included in two multicenter studies primarily focusing oncirculating tumor cell detection (study 06-248 [26, 27] and study 09-405 [28]). Patients were recruited between February 2008 and March 2015 in several hospitals in the Netherlands and Belgium. The study protocols were approved by the local research ethics boards (Erasmus MC identifiers MEC-06- 248 and MEC-09-405) and conducted in accordance with the Declaration of Helsinki. All patients gave written informed consent before participation.

A total cohort of 156 MBC patients with at least 200 µL plasma available was evaluated; 113 patients were treated with endocrine therapy, and 43 patients were treated with chemotherapy, for MBC. More detailed information has been summarized in **Table 1** for the 95 clinically evaluable endocrine-treated patients and in **Table S1** for all patients whose plasma was used in this study. In addition, plasma samples from 18 healthy blood donors (HBDs) were assessed.

Plasma collection and cfDNA isolation

Blood (10 mL) was collected in Vacutainer® EDTA tubes (BD, Franklin Lakes, NJ, USA) and transported to the coordinating lab at ambient temperature for CTC profiling and plasma preparation, both processed within 24 hours after blood draw. Of this blood, 7.5 mL was used for circulating tumor cell enrichment and characterization by quantitative RT-PCR (RT-qPCR) as described before [26, 29] and the remaining up to 2.5 mL whole blood to prepare plasma. The plasma was collected after two sequential centrifugation steps (1,711*g for 10 minutes at room temperature and immediately stored at -80°C and a second time at 12,000*g for 10 minutes at 4°C before cfDNA isolation).

The cfDNA extraction was performed using the QIAamp® Circulating Nucleic Acid kit (Qiagen, Venlo, The Netherlands) according to the manufacturer's protocol with some modifications: i) dithiothreitol (DTT) at a 5 mM/L final concentration was added to the plasma before thawing to prevent nucleic acid degradation; ii) the input volume of plasma was downscaled to 200 µL and iii) the cfDNA was eluted in 20 µL AVE-Buffer (Qiagen, Venlo, The Netherlands) and applied to the column three times to increase the final cfDNA concentration. The cfDNA concentration was quantified with the Qubit dsDNA HS Assay kit (Thermo Fisher, Landsmeer, The Netherlands) and stored for use within one week at -30°C and long term at -80°C.

Table 1. Treatment characteristics, endocrine clinical cohort

Parameter	Baseline subset (n= 51)	PD subset (n= 44)
Age at sample draw Median age (range)	67 (36-89)	67 (39-87)
Adjuvant endocrine therapy		
No	33 (65%)	27 (61%)
Yes, tamoxifen only	12 (23%)	10 (23%)
Yes, tamoxifen + AI	2 (4%)	6 (14%)
Yes, AI only	4 (8%)	1 (2%)
Adjuvant chemotherapy		
No	38 (75%)	32 (73%)
Yes	13 (25%)	12 (27%)
Neoadjuvant therapies		
No	51 (100%)	42 (95%)
Yes, chemotherapy		2 (5%)
Previous endocrine therapy lines for MBC, *n*		
0	51 (100%)	
1		28 (64%)
2		13 (29%)
≥3		3 (7%)
Endocrine therapy after start (BL subset) or before PD (PD subset)		
AI	37 (73%)	22 (50%)
Tamoxifen	13 (25%)	11 (25%)
Fulvestrant	1 (2%)	10 (23%)
Fulvestrant + AI		1 (2%)
Previous endocrine therapy lines for MBC (in case of inclusion at PD on ≥2nd-line endocrine therapy)		
No	51 (100%)	28 (64%)
Yes, AI only		10 (23%)
Yes, AI + tamoxifen		3 (7%)
Yes, tamoxifen only		3 (7%)

Data are given as *n* (%) of each group, unless otherwise indicated.

AI, aromatase inhibitor; BL, baseline; MBC, metastatic breast cancer; PD, progressive disease.

Pre-Amplification and Quantitative PCR

cfDNA (0.1 to 1 ng/ µL) was pre-amplified during 15 cycles using a single locus (specific for *ESR1* containing the 4 hotspot mutations at codon positions 537 and 538) or a multiple loci target-specific amplification for *ESR1, BRAF*, *KRAS*, *PIK3CA,* and *TP53* with TaqMan PreAmp Master Mix (Thermo Fisher) as recommended by the manufacturer. All primer pairs present in the two different pre-amplification combinations, as well as the Taqman qPCR and digital PCR assays used to quantify the wild-type and mutated genes, are given in **Table 2;** details regarding the pre-amplification and digital PCR protocols in **Table S2**. Before downstream processing, the pre-amplified product was diluted 10-fold in LoTE buffer (3 mM Tris-HCl/0.2 mM EDTA, pH 8.0).
Next, a quantitative PCR (qPCR) for wild type (WT) *ESR1* was performed to quantify the number of wild-type (WT) copies present in the samples. To ensure equal WT copies loading of all samples, the resulting quantification cycle (Cq) value was used to empirically establish the correct dilution factor to apply to each sample before loading onto the chips for the dPCR analysis (**Table S3**).

Chip-based digital PCR

dPCRs were performed with the QuantStudio 3D Digital PCR System (Thermo Fisher) according to the manufacturer's protocol. All four hotspot mutations for *ESR1* (D538G, Y537S, Y537C and Y537N) were separately, or together by multiplexing the assays, analyzed with mutation-specific Taqman assays (**Table 2** and **Table S2**) and the variant allele frequencies (VAFs) were calculated for each of these mutations or for the combined mutations, respectively. For this, each pre-amplified cfDNA sample was partitioned into 20,000 wells of a QuantStudio 3D Digital PCR v2 Chip and run on a ProFlex 2x Flat PCR System (Thermo Fisher). The target-specific optimized PCR program was: 10 min at 96°C, followed by 40 cycles of 30 s incubation at 98°C and 2 min at 55°C and a final pause for up to 16 hours at 10°C. Chips were read in a QuantStudio 3D Digital PCR Instrument and analyzed with the web-based Quantstudio 3D dPCR Analysis Software version 3.01 (Thermo Fisher). At least one positive and one negative control sample was included in every run, and all chips were read in duplicate (immediately after the dPCR run and again between 4 to 16 hours later).

Table 2. Assay details pre-amplification and digital PCR

Description	Mutation	Assay ID*	Forward primer	Reverse primer	FAM Taqman probe **	VIC Taqman probe **	Size, bp
***ESR1* uniplex model**							
ESR1 PreAmp[6]	*ESR1*		5′-AGGCATGGAG CATCTGTACA-3′	5′-TTGGTCC GTCTCCTCCA-3′			136
ESR1 mutation	*ESR1_D538G*		5′-CAGCATGAAG TGCAAGAACGT-3′	5′-TGGGCGTCC AGCATCTC-3′	5′-CCCTCTATG GCCTGCT-3′	5′-CCCCTCTA TGACCTGCT-3′	63
ESR1 mutation	*ESR1_Y537C*[6]		5-AGGCATGGA GCATCTGTACA-3′	5′-TTGGTCCG TCTCCTCCA-3′	5′TGCCCCT CTGTGACCTGC-3′	5′-TGGTGCCCCT CTATGACCTG-3′	136
ESR1 mutation	*ESR1_Y537N*[6]		5′-AGGCATGGAG CATCTGTACA-3′	5′-TTGGTCCG TCTCCTCCA-3′	5′-TGCCCCTCA ATGACCTGC-3′	5′-TGGTGCCCCT CTATGACCTG-3‘	136
ESR1 mutation and qPCR screening	*ESR1_Y537S*		5′-CAGCATGAAGTGC AAGAACGT-3′	5′-TGGGCGTC CAGCATCTC-3′	5′-CCCTCTCT GACCTGC-3′	5′-CCCCTCT ATGACCTGC-3′	63
Additional assays included in the multiplex model							
KRAS PreAmp	*KRAS_codon 12/13*		5′-ACTGGTGGAGTA TTTGATAGTGTAT-3′	5′-CTCTATTGTTGGA TCATATTCGTCC-3′			204
BRAF PreAmp	*BRAF_V600E*	AH6R5PH					149
PIK3CA PreAmp	*PIK3CA_E542K_760*	AHD2BSD					
PIK3CA PreAmp	*PIK3CA_H1047L_776*	AHLJ0TP					
TP53 PreAmp	*TP53_Exon 5*		5′-CCCCTGCCCTCA ACAAGATG-3′	5′-GACCATCGCTA TCTGAGCAG-3′			183
TP53 PreAmp	*TP53_Exon 7*		5′-TGGCTCTGACTG TACCACCA-3′	5′-CTGGAGTCTTC CAGTGTGATG-3′			108
TP53 PreAmp	*TP53_Exon 8*		5′-ACTGGGACG GAACAGCTTTG-3′	5′-CTGGGGGCAGCTCGTG-3′			113
TP53 PreAmp	*TP53_Exon 10*		5′-GTGAGCGCTTC GAGATGTTC-3′	5′-TCCCCCCTGGCTCCTTC-3′			81
Reference; synthetic *ESR1*_D538G		5′-GTC-TTC-CCA-CCT-ACA-GTA-ACA-AAG-GCA-TGG-AGC-ATC-TGT-ACA-GCA-TGA-AGT-GCA-AGA-ACG-TGG-TGC-CCC-TCT-ATG-GCC-TGC-TGC-TGG-AGA-TGC-TGG-ACG-CCC-ACC-GCC-TAC-ATG-CGC-CCA-CTA-GCC-GTG-GAG-GGG-CAT-CCG-TGG-AGG-AGA-CGG-ACC-AAA-GCC-ACT-TGG-CCA-CTG-CGG-GCTC-3′					181
Reference; synthetic *ESR1*_Y537C [6]		5′-GTC-TTC-CCA-CCT-ACA-GTA-ACA-AAG-GCA-TGG-AGC-ATC-TGT-ACA-GCA-TGA-AGT-GCA-AGA-ACG-TGG-TGC-CCC-TCT-GTG-ACC-TGC-TGC-TGG-AGA-TGC-TGG-ACG-CCC-ACC-GCC-TAC-ATG-CGC-CCA-CTA-GCC-GTG-GAG-GGG-CAT-CCG-TGG-AGG-AGA-CGG-ACC-AAA-GCC-ACT-TGG-CCA-CTG-CGG-GCTC-3′					181

Reference; synthetic *ESR1*_Y537N [6]	5’-GTC-TTC-CCA-CCT-ACA-GTA-ACA-AAG-GCA-TGG-AGC-ATC-TGT-ACA-GCA-TGA-AGT-GCA-AGA-ACG-TGG-TGC-CCC-TCA-ATG-ACC-TGC-TGC-TGG-AGA-TGC-TGG-ACG-CCC-ACC-GCC-TAC-ATG-CGC-CCA-CTA-GCC-GTG-GAG-GGG-CAT-CCG-TGG-AGG-AGA-CGG-ACC-AAA-GCC-ACT-TGG-CCA-CTG-CGG-GCTC-3’	181
Reference; synthetic *ESR1*_Y537S [6]	5’-GTC-TTC-CCA-CCT-ACA-GTA-ACA-AAG-GCA-TGG-AGC-ATC-TGT-ACA-GCA-TGA-AGT-GCA-AGA-ACG-TGG-TGC-CCC-TCT-CTG-ACC-TGC-TGC-TGG-AGA-TGC-TGG-ACG-CCC-ACC-GCC-TAC-ATG-CGC-CCACTA-GCC-GTG-GAG-GGG-CAT-CCG-TGG-AGG-AGA-CGG-ACC-AAA-GCC-ACT-TGG-CCA-CTG-CGG-GCTC-3’	181

*Thermo Fisher; **MGB NFQ probe. FAM, 6-carboxyfluorescein; MGB NFQ, Minor Groove Binder NonFluorescent Quencher; qPCR, quantitative PCR.

Ion Torrent Next-Generation Sequencing

In addition to dPCR, cfDNA of 10 HBDs as controls for accuracy and specificity and serial samples of 7 MBC patients were analyzed using the Ion Torrent™ Oncomine™ Breast cfDNA Assay (Catalog Number: A31183, Thermo Fisher) in combination with the Ion Torrent S5XL NGS system to simultaneously evaluate multiple hotspot mutations. The Oncomine Breast Assay is designed to sequence 26 amplicons to detect 152 hotspots and indels for a panel of 10 BC-relevant genes (*AKT1*, *EGFR*, *ERBB2*, *ERBB3*, *ESR1*, *FBXW7*, *KRAS*, *PIK3CA*, *SF3B1* and *TP53*). This NGS Assay applies unique molecule identifiers to improve the sensitivity by decreasing the amount of sequencing artefacts. With the recommended input of 20 ng cfDNA, the use of unique molecule identifiers enables a limit of detection (LOD) as low as 0.1% [30-33].

Because of our retrospective analysis, suboptimal input amounts compared to the recommended input of 20 ng DNA were used for cfDNA from HBDs (range 4.86 to 10.41 ng) and MBC patients (range, < 1 to 6.1 ng) to generate targeted libraries following the manufacturer's protocol. First, concentrations of each Oncomine™ cfDNA library were determined by qPCR using the Ion Library TaqMan® Quantitation Kit (Life Technologies) and then diluted to a final concentration of 50 pM/L. Next, sample barcoded libraries were pooled together for template preparation on the Ion Chef™ Instrument using the Ion 520/530™ Kit – Chef (Catalog number: A30010) and loaded onto an Ion 530™ chip. The chip was sequenced on an Ion S5™ XL Sequencer Systems and the data were analyzed using the Ion Torrent Suite™ Software 5.2.2. For patient and HBD samples, general NGS quality measures (eg, median read depth, median molecular coverage, and mean read lengths) are presented in **Table S4.**

Statistical analysis

All data were entered in SPSS version 23 (IBM Corp., Armonk, NY, USA) to generate the tables and perform the statistical analyses. For 2x3 and 2x4 tables, the online Freeman-Halton extension of the Fisher's exact probability test was used (http://vassarstats.net/newcs.html; date of last access October 1, 2018). All *P*-values are two-sided and $P<0.05$ was considered statistically significant.

Results

Sensitivity and specificity of a multiplex mutant *ESR1* hotspot dPCR screening assay

We designed a multiplex dPCR *mESR1* screening assay (m*ESR1*_G_C_N_S) covering the four most common *ESR1* mutations (D538G, Y537C, Y537N and Y537S) (**Table 2** and **Table S2**). To validate the performance of our m*ESR1*_G_C_N_S screening assay, 36 plasma samples were first analyzed with the m*ESR1* dPCR screening assay and subsequently compared with the four individual *ESR1* mutation assays (**Table S5**, 'Experiment description' dPCR multiplex screening). Using a VAF threshold of 2% for the m*ESR1* screening assay, 12 plasma samples were above the threshold for the multiplex assay (VAF range 2.04% to 27.73%). Digital PCR analysis using individual mutation assays demonstrated that at this cut-off all individual hotspot mutations using a VAF threshold of 1% for the uniplex assays were efficiently captured with the m*ESR1* hotspot dPCR screening assay. Only one sample (Sample 70) with an *ESR1* Y537C mutation at 1.22% was missed with the dPCR m*ESR1* screening assay giving a VAF of 0.43% (**Table S5**, column *ESR1*_G_C_N_S). **Figure 2** shows that the sum of the frequencies of the individual Taqman assays correlated well with the VAF assessed with the m*ESR1* dPCR screening assay (Pearson's $r = 0.99$, $P<0.0001$). In addition, as shown for Samples 9 and 25, the multiplex screening assay was also able to capture the possible presence of multiple mutations in the same sample, suggesting a different clonal background. The m*ESR1*_G_C_N_S dPCR screening approach was, therefore, applied to screen our samples before performing the individual dPCR *ESR1* mutation assays.

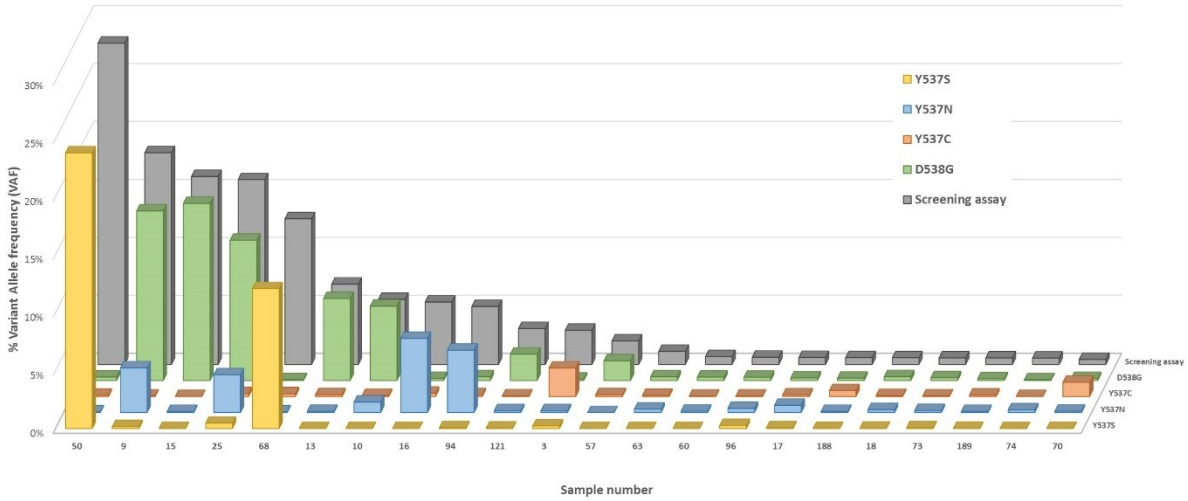


Figure 2: Performance m*ESR1* dPCR screening assay.

To validate the performance of the m*ESR1* multiplex screening assay covering all four *ESR1* hotspot mutations, 36 different plasma samples were analyzed with the m*ESR*1 dPCR screening assay (grey bars). Variant allele frequency data were compared with the results of the individual m*ESR1* assays (Y527S; yellow bar, Y537N; blue bars, Y537C; orange bars and D538G, green bars).

Sensitivity and specificity of single versus multilocus pre-amplification

Because limited amounts of plasma were available, pre-amplification was required to assay multiple mutations in the same sample. Because whole genome amplification using, for example, the REPLI-g Whole Genome Amplification method (Qiagen) is not suitable for fragmented (cf)DNA and, therefore, failed for our targets (data not shown), a target-specific pre-amplification method was designed to allow measurement of *ESR1* mutations next to *KRAS*, *BRAF*, *PIK3CA* and *TP53* in cfDNA isolated from down to 200 µL plasma. These analyses showed that, using a 1% VAF threshold for the 4 individual mutations, no information was lost if the *ESR1* locus was pre-amplified in a multiplex protocol together with *KRAS*, *BRAF*, *PIK3CA* and *TP53* (**Table S5**, 'Experiment description' preamp uni_multi at 2 µL). All samples identified as mutated at a 1% VAF threshold with the *ESR1* target specific pre-amplification protocol (**Table S5**) were also identified as such with the multiplex pre-amplification protocol (**Table S5**).

Range of plasma volume input for extraction, DNA input for digital PCR and cfDNA input for pre-amplification

To determine the minimum volume of plasma required for our analyses, cfDNA was isolated from different volumes of plasma from the same HBD with and without spiked-in DNA containing

D538G *ESR1* (isolated from a D538G positive formalin-fixed, paraffin-embedded tissue) and then analyzed by dPCR both with and without *ESR1* locus-specific pre-amplification. An average VAF of 20.4% (range, 14.6% to 28.3%) and 18.0% (range, 15.3% to 25.8%) for D538G *ESR1* was measured in unprocessed and pre-amplified samples, respectively (**Table 3**), indicating that similar allele frequencies were detected in cfDNA isolated from 4 mL down to 200 µL of plasma with and without pre-amplification. Similar starting volumes of HBD plasma without spiked-in cfDNA were used as negative control. While no D538G *ESR1* copies were detected in non-pre-amplified HBD samples, we measured an average VAF of 0.3% (range, 0.2% to 0.5%) in pre-amplified samples, which is, however, well below the 1% cut-off we used to assign a sample mutated (see also below). In conclusion, an input of 200 µL plasma was considered sufficient as input to detect variant molecules in cfDNA after pre-amplification.

Table 3. Sensitivity and Specificity of Measuring D538G ESR1 Variant Allele frequencies

Sample		Without pre-amplification			With pre-amplification		
HBD plasma, mL	**D538G spiked-in**	**dPCR, pg***	**VAF%**	**VAF, means ± SD, %**	**dPCR, pg***	**VAF%**	**VAF, means ± SD%**
4.0	Yes	11955	18.4		9444	16.1	
2.0	Yes	7601	14.6		7261	17.4	
1.0	Yes	4543	20.2		8806	15.3	
0.5	Yes	2119	28.3		7489	15.7	
0.2	Yes	2798	20.6	20.4 ± 5.0	6891	25.8	18.0 ± 4.4
2.0	No	1999*	0,0		11965	0.5	
1.0	No	1182*	0,0		7968	0.2	
0.2	No	412*	0,0	0.0 ± 0.0	12076	0.2	0.3 ± 0.2

*Less than 2000 pg DNA input in the dPCR is considered a too low input for a reliable analysis of low mutated samples. dPCR, digital PCR; HBD, Healthy blood donor; VAF, variant allele frequency.

Because the estimated VAF in dPCR at a final input of < 2,000 pg DNA (17.4% ± 4.3% (25% CV)) was lower and less reliable compared with a VAF assessed in at least 2,000 pg DNA (22.2% ± 1.0% (4.5% CV)) (**Table 4**), it was decided that at least 2,000 pg cfDNA input was required for dPCR. On the other hand, to prevent overloading of the dPCR chip, the most optimal maximum input for the dPCR needed to be determined. The VIC signal of our qPCR for Y537S *ESR1* (with the same Taqman assay used for the dPCR) was used to assess the number of WT *ESR1* copies present in the pre-amplified cfDNA, irrespective of the mutation present. Using the resulting quantification cycle (Cq) value as an estimate for the number of WT copies, the input range for the m*ESR1* dPCR

was empirically established after pre-amplification to load the chip with between 2 and 100 ng WT *ESR1* DNA (**Table S3**). When using this approach, the Cq values versus input range of pre-amplified material need to be optimized first for any other dPCR assay.

Table 4. dPCR input evaluation for measuring D538G *ESR1* variant allele frequencies

Without pre-amplification		**With pre-amplification**		
dPCR, pg*	**VAF, %**	**dPCR, pg**	**VAF, %**	**VAF, means ± SD %**
116*	14.8	641*	18.4	
		1227*	25.5	
		1642*	14.8	
		1722*	15.9	
		1770*	16.6	
		1786*	13.3	**17.4 ± 4.3**
		2756	21.9	
		7777	24.1	
		12254	23.3	
		12371	21.6	
		17216	22.7	
		18147	22.4	
		20118	21.1	
		20125	21.1	
		24859	22.2	
		309706	21.5	**22.2 ± 1.0**

Less than 2000 pg DNA input in the dPCR is considered a too low input for a reliable analysis of low mutated samples. dPCR, digital PCR; HBD, Healthy blood donor; VAF, variant allele frequency.

To determine if the amount of cfDNA input in the pre-amplification biased the resulting VAF, cfDNA isolated from plasma of 26 patients was pre-amplified using 0.5 µL cfDNA [2.5% of the total isolated volume, median (range): 0.28 (0.17 to 3.9) ng] in a uniplex pre-amplification assay and 2 µL [10%, median (range): 1.12 (0.68 to 15.6) ng)] in a multiplex pre-amplification assay. Similar VAFs were observed for 24 of the 26 patients (**Table 5**). *ESR1* mutations were detected in eight samples pre-amplified using 0.5 µL cfDNA input (7x D538G, 1x Y537C and 1x Y537N) and in nine samples using 2 µL input (7x D538G, 2x Y537C and 2x Y537N)**.** According to both pre-amplification protocols one sample (Sample 9) showed 2 different mutations (D538G and Y537C). However, after using the multiplex pre-amplification protocol with 2 µL of cfDNA, the presence of an additional mutation in Y537N (6.3% VAF) was identified in Sample 25, which was not detected using the uniplex pre-amplification protocol with 0.5 µL cfDNA input (**Table 5,** left

columns). Thus, increasing the cfDNA input for the pre-amplification to 2 µL led to the detection of one additional mutated sample. Because the volume of cfDNA in the pre-amplification can affect downstream analyses, additional tests were performed using higher cfDNA input volumes of 8 to 10 µL plasma (40% to 50% of the cfDNA isolated from 200 µL plasma). The data obtained in these analyses were similar to the data obtained with 2 µL cfDNA (**Table 5**, right columns) indicating that 2 µL cfDNA input in the pre-amplification was sufficient to identify all four *ESR1* mutations higher than the cutoff level of 1%.

Table 5. Sensitivity and specificity of measuring *ESR1* variant allele frequencies in pre-amplified cf-DNA

	UNIPLEX pre-amplification (0.5 µL cfDNA)				**MULTIPLEX pre-amplification (2 µL cfDNA)**				**MULTIPLEX pre-amplification (10 µL vacum-dried cfDNA)**				
Sample ID	**D538G**	**Y537C**	**Y537N**	**Y537S**	**D538G**	**Y537C**	**Y537N**	**Y537S**	**G_C_N_S**	**D538G**	**Y537C**	**Y537N**	**Y537S**
1	0.2	0.1	0.0	0.0	0.0	0.2	0.1	0.0	0.5	0.3	0.0	0.0	0.0
2	0.1	0.1	0.1	0.0	0.3	0.2	0.2	0.0					
3	0.2	0.0	0.1	0.1	0.1	**2.5**	0.2	0.2					
4	0.1	0.3	0.0	0.0	0.1	0.1	0.1	0.2					
5	0.4	0.4	0.2	0.0	0.2	0.4	0.2	0.0					
6	0.1	0.1	0.4	0.0	0.2	0.4	0.2	0.1	0.4				
7	0.3	0.2	0.1	0.0	0.2	0.1	0.1	0.0	0.2				
8	0.5	0.2	0.2	0.1	0.3	0.1	0.1	0.0	0.2				
9	**27.1**	**12.1**	0.1	0.0	**14.6**	**3.9**	0.0	0.2					
10	**2.9**	0.2	0.3	0.0	**6.0**	0.4	0.7	0.1	**5.6**	**6.4**	0.1	0.9	0.0
11	**11.3**	0.4	0.2	0.0	**11.0**	0.5	0.7	0.0					
12	0.2	0.2	0.1	0.1	0.5	0.1	0.1	0.0	0.1				
13	**5.1**	0.1	0.0	0.1	**5.1**	0.2	0.4	0.0	**6.9**	**7.1**	0.2	0.1	0.0
14	0.7	0.2	0.1	0.1	0.3	0.1	0.3	0.5	0.4				
15	**10.2**	0.1	0.3	0.1	**17.8**	0.3	0.2	0.0	**16.2**	**15.3**	0.1	0.0	0.0
16	0.2	0.1	**6.3**	0.0	0.2	0.3	**6.4**	0.0					
17	0.5	0.1	0.4	0.1	0.2	0.1	0.5	0.6	0.6	0.2	0.6	0.1	0.1
18	0.2	0.2	0.1	0.0	0.4	0.6	0.3	0.0	0.6	0.3	0.3	0.1	0.0
19	0.3	0.4	0.1	0.1	0.2	0.2	0.1	0.0	0.3				
20	0.1	0.3	0.1	0.0	0.5	0.2	0.1	0.0	0.1				
21	0.3	0.2	0.0	0.1	0.3	0.2	0.0	0.0	0.1				
22	0.1	0.1	0.5	0.1	0.9	0.1	0.1	0.0	0.1				
23	0.1	0.0	0.2	0.1	0.0	0.2	0.0	0.0	0.1				
24	**20.6**	0.2	0.2	0.1	**20.5**	0.0	0.1	0.0					
25	**16.0**	0.1	0.2	0.0	**8.7**	0.1	**6.3**	0.0	**16.0**	**12.1**	0.2	**3.2**	0.5
26	0.4	0.2	0.1	0.5	0.6	0.1	0.1	0.0	0.2				

Data are given as variant allele frequency. Values > 1% threshold cutoff are in bold.
cfDNA, cell-fre DNA.

Reproducibility of VAF estimates after pre-amplification

To evaluate the reproducibility of our method to analyze mutations in minute amounts of cfDNA, the four hotspot *ESR1* mutations were measured by performing multiple independent technical replicates using a pool of cfDNA (multisample pool) containing mutated copies of three of our four *ESR1* hotspot mutations. D538G *ESR1* was not included and, thus, served as a negative control. The VAFs for the individual mutations present in the pooled sample were analyzed after independent pre-amplification sessions with both the uniplex and multiplex pre-amplification protocol. The independent replicates resulted in similar VAFs for both the protocols (**Table 6**). In addition, a frequency of 50.1% ± 0.8% was detected when the *ESR1* dPCR prescreening assay containing all four individual Taqman assays was used. As observed before, this frequency equals the sum of the VAFs (56.9% after uniplex and 51.2% after multiplex pre-amplification) of the four individual *ESR1* mutations.

Table 6: Reproducibility of VAF estimates after pre-amplification of a multipool sample

Pre-amplification protocol	*ESR1* mutation	pg in dPCR	VAF, %	VAF, means ± SD%
ESR1 UNIPLEX	D538G	23742	0.30	
ESR1 UNIPLEX	D538G	23881	0.30	
ESR1 UNIPLEX	D538G	31284	0.30	**0.32 ± 0.00**
MULTIPLEX	D538G	3206	0.20	
MULTIPLEX	D538G	3826	0.40	
MULTIPLEX	D538G	3851	0.50	
MULTIPLEX	D538G	35892	0.25	**0.34 ± 0.14**
ESR1 UNIPLEX	Y537C	25759	38.05	
ESR1 UNIPLEX	Y537C	34931	40.25	
ESR1 UNIPLEX	Y537C	58852	35.20	**37.83 ± 2.53**
MULTIPLEX	Y537C	6540	32.10	
MULTIPLEX	Y537C	6559	32.15	
MULTIPLEX	Y537C	28689	36.10	**33.45 ± 2.30**
ESR1 UNIPLEX	Y537N	17949	14.65	
ESR1 UNIPLEX	Y537N	26306	13.80	
ESR1 UNIPLEX	Y537N	45277	13.80	**14.08 ± 0.49**
MULTIPLEX	Y537N	4806	12.60	
MULTIPLEX	Y537N	4908	12.35	
MULTIPLEX	Y537N	20837	12.30	**12.42 ± 0.16**
ESR1 UNIPLEX	Y537S	27800	4.95	
ESR1 UNIPLEX	Y537S	27801	4.10	
ESR1 UNIPLEX	Y537S	29441	4.95	**4.67 ± 0.49**
MULTIPLEX	Y537S	3951	2.90	
MULTIPLEX	Y537S	4496	5.90	
MULTIPLEX	Y537S	4663	6.00	
MULTIPLEX	Y537S	34605	3.25	**5.05 ± 1.56**
ESR1 UNIPLEX	*ESR1*_G_S_C_N	59887	50.70	
ESR1 UNIPLEX	*ESR1*_G_S_C_N	56253	49.55	**50.13 ± 0.81**

dPCR, digital PCR; VAF, variant allele frequency.

Genomic DNA isolated from the m*ESR1* negative MDA-MB-435s cell line was also fragmented by sonication to obtain a negative control sample with the same characteristics of cfDNA (which is generally fragmented). Measuring the *ESR1* mutational status after eight independent pre-amplification experiments with our multiplex pre-amplification protocol showed that, also for this negative control sample, the results were reproducible and compared well with the nonamplified parental cell line (**Table S5**, 'Experiment description' Reproducibility fragmented cell line). In summary, the following are results for the parental cells before pre-amplification versus the data obtained after the eight independent multiplex pre-amplifications: D538G, 0.00% versus 0.24% ± 0.15%; Y537C, 0.45% versus 0.17% ± 0.19%; Y537N, 0.47% versus 0.22% ± 0.24%; Y537S, 0.00% versus 0.02% ± 0.03%, and for the screening multiplex dPCR assay, 0.20% vs 0.51% ± 0.22%.

Lower limit of detection

Now that we optimized the cfDNA input volume for pre-amplification at 2 µL, a cutoff could be set up for our assays to assign a patient sample mutated for *ESR1.* The lowest VAF that could distinguish a negative VAF at three times the baseline noise was first estabilished. For D538G, the LOD was calculated to be 0.89% in our multiplex pre-amplification model (**Table S6**). In addition, the four *ESR1* mutations were measured in 18 individual HBDs and the highest VAF + 2.58 SD was calculated to achieve a LOD with 99% confidence. These analyses resulted in the following cutoffs: 1.02% for D538G, 0.52% for Y537C, 0.87% for Y537N and 0.97% for Y537S (**Table S7**). For our patient samples, it was, therefore, decided to use a cutoff of at least 1% mutated for any of the hotspot mutations before a patient could be called positively mutated for *ESR1*.

Comparison with Next Generation Sequencing

To explore whether targeted mutation analysis by NGS would confirm and/or improve sensitivity to detect circulating tumor DNA by m*ESR1* compared to our digital PCR m*ESR1* workflow, cfDNA samples of seven MBC patients at baseline, while receiving treatment and/or at progressive disease were evaluated by both techniques.

A total of 17 cfDNA samples from seven MBC patients and 10 cfDNA samples from 10 HBDs as controls were analyzed by means of the Ion Torrent™ Oncomine™ Breast cfDNA Assay. The NGS results were defined as successful in this study when the following occurred: i) samples had a median number of > 100 independent molecules sequenced (**Figure 3A**), ii) the variants were detected in at least two independent molecules, and iii) the VAF was higher than the calculated LOD of the NGS assay. No variants were detected in any of the control samples and in 7 of 17 patient samples in the 10 genes analyzed by the targeted Oncomine NGS panel. This failure to detect variants and the low-molecule coverage were explained by cfDNA input amounts of < 5 ng in eight case samples (**Table S4**). The remaining nine case samples had a total of 22 variants, with 10 Catalogue of Somatic Mutations in Cancer (COSMIC)-reported hotspot missense mutations in four patients (44%) and detected in the genes *ESR1* (D538G, Y537N, Y537S,and E380Q), P*IK3CA* (H1047R, E726K), *KRAS* (G12C), and *TP53* (R213L, V274F,and R280T). All variants detected in cases were not present in the controls, although the positions were sequenced at similar or even higher median molecule coverage.

For all the seven MBC patients, the *ESR1* mutational status was investigated by both NGS and dPCR. Although the measured VAFs differed slightly by the two methods, four of the five *ESR1* mutations detected by dPCR were retrieved by NGS (**Figure 3B**). Only the *ESR1* D538G measured in a sample of case 60 by dPCR (VAF, 1.90%), was missed by the NGS approach. The latter may be explained by the differences in cfDNA input (i.e. 18 ng of pre-amplified cfDNA as input for dPCR and 4.3 ng cfDNA as input for NGS). The NGS analysis discovered the *ESR1* E380Q mutation, albeit at low frequency (VAF, 0.48%), in a progressive disease sample of Case 77. However, this mutation was not included in our *ESR1* pre-amplification mix. Finally, Cases 16 (**Figure 3C**) and 77 (**Figure 3D**) had, next to the *ESR1* mutation found at progressive disease, also *PIK3CA* mutations detected at baseline and at progressive disease.

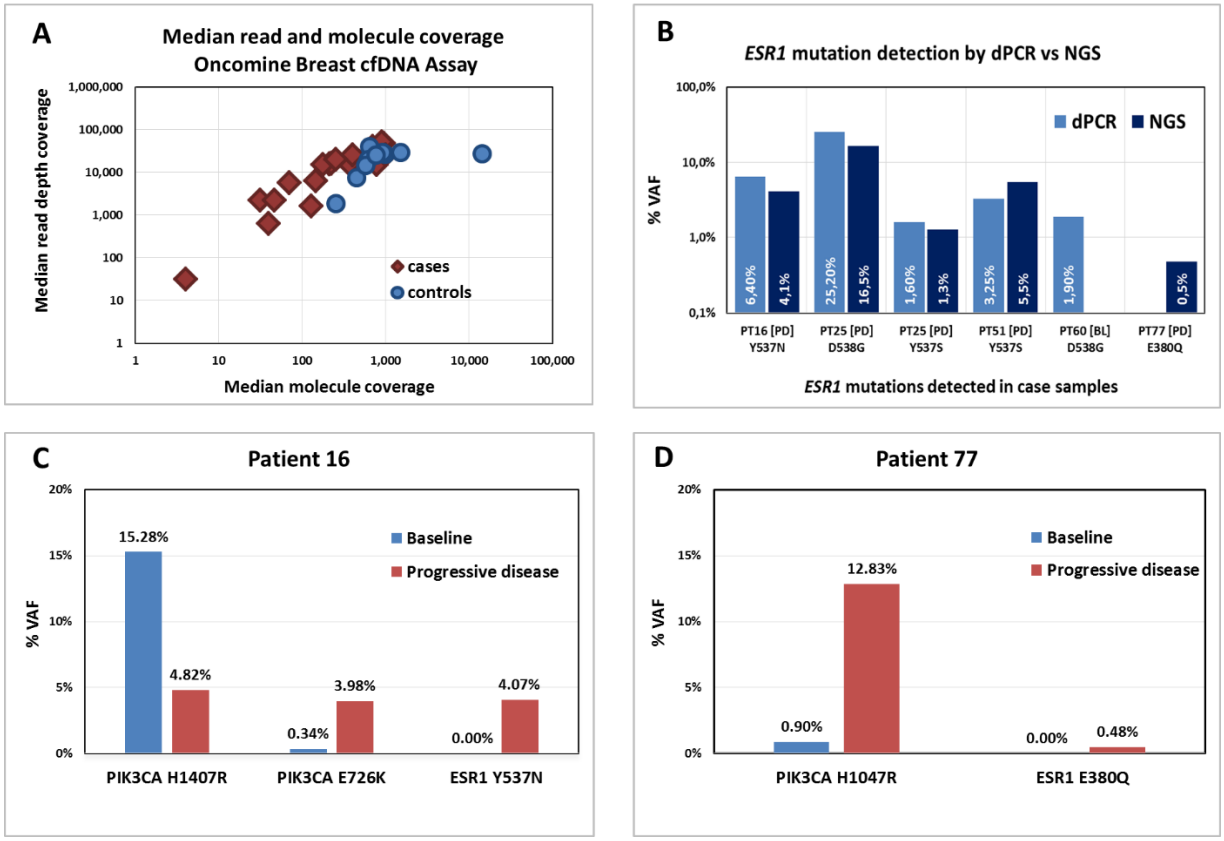


Figure 3. Digital PCR (dPCR) *versus* targeted next-generation sequencing (NGS) with Oncomine Breast cfDNA Assay.

A) The median molecule coverage versus the median read depth coverage of the investigated cases (red diamonds) and controls (blue circles) by NGS. Five cases did not pass the cutoff of > 100 independent molecules sequenced. **B)** Comparison of *ESR1* mutation status evaluated by dPCR (light blue bars) and NGS (dark blue bars) for five samples for which both dPCR and NGS data were available. Only the *ESR1* D538G, measured in a sample of Case 60 by dPCR [variant allele frequency (VAF, 1.93%)], was missed by the NGS approach. The low-frequency E380Q detected by NGS was not tested by dPCR**. C)** and **D)** Different *ESR1* and *PIK3CA* mutations detected by NGS in matched baseline (BL) (blue bars) and progressive disease (PD) (red bars) samples of two cases (Case 16 and 77).

ESR1 mutations in clinical cohorts

The optimized workflow was applied on a total of 156 samples of 132 individual patients who received first (67 patients; 81 samples) and/or later lines of endocrine treatment (35 patients; 35 samples) or first line chemotherapy (37 patients; 40 samples) (**Table S1**).

The *ESR1* mutational status of patients receiving first-line endocrine therapy for MBC (baseline subset, *n*=51) was first compared with that of patients who had progressed on endocrine therapy (progressing subset, *n*=44). The patient characteristics, including endocrine therapies received before the plasma sampling, are summarized in **Table 1.** In the progressive subset, a higher frequency of cases with *ESR1* mutations (34.1%) was observed compared with the baseline subset (3.9%) (Fisher's exact test *P* = 0.0001) (**Table 7**).

Table 7: *ESR1* mutational status in endocrine treated patients according time of plasma sampling

***ESR1* mutated at 1% cut-off**	**Baseline**		**Progressive Disease**		***P* value (Fisher's Exact Test)**
	count	**%**	**count**	**%**	
No	49	96.1	29	65.9	0.0001
Yes	2	3.9	15	34.1	

Next, the *ESR1* mutational status for the chemotherapy and endocrine cohorts was analyzed separately, subdivided in baseline samples taken before start of first line therapy, samples taken at progressive disease, and samples taken in between these two time-points (on-treatment phase, eg. 1 week up to 6 months from the start of any line of therapy). These analyses revealed that, although numbers for the chemotherapy cohort at the on-treatment and progressive disease stage were small, similar percentages of *ESR1* mutated samples were found at baseline in the chemotherapy and endocrine therapy cohorts (**Table 8**).

Table 8: ESR1 mutational status according inclusion cohort

Therapy started after inclusion	***ESR1* mutated at 1% cutoff**	**Chemotherapy (*n* = 43)**		**Endocrine therapy (*n* = 113)**	
		n	%	*n*	%
Baseline	no	15	34.9	49	43.4
	Yes	1	2.3	2	1.8
On treatment	No	21	48.8	20	17.7
	Yes	1	2.3	0	0.0
Progressive disease	No	4	9.3	27	23.9
	Yes	1	2.3	15	13.3

To further evaluate the prognostic and predictive value of liquid biopsies, m*ESR1* was analized in longitudinally collected cfDNA of 17 endocrine-treated MBC patients (**Table S8**). A minimum of two and up to a maximum of five follow-up samples were available for each patient. Thirteen patients lacked m*ESR1* during follow-up. One patient (Patient 60) had an m*ESR1* (1.90% for D538G) at baseline which was no longer detected at 1 and 9 months after tamoxifen, although an increase (from 0.5 to 1.3 ng/uL) in cfDNA concentration was observed in the first follow-up sample, followed by a decrease down to baseline level 8 months later (**Figure 4A**). The patient was still alive without progression 2 years later at the last recorded follow-up data. In three additional patients (Patients 16, 25 and 51), an acquired m*ESR1* reflected the putative presence of progressive disease. Patient 16 acquired a Y537N mutation (from 0.02% to 6.39%) after letrozole, and Patient 51 acquired a Y537S mutation (from 0.00% to 3.25%) after tamoxifen treatment (**Table S8**). Patient 25 received exemestane as first line therapy. At baseline and in two follow-up samples during the aromatase inhibitor treatment, no m*ESR1* was detected. However, at progression a high frequency of the D538G mutation (25.16%) and a lower frequency of the Y537N mutation (1.61%) were detected in cfDNA, reflecting disease progression with a likely polyclonal mutation status. Fulvestrant plus dovitinib in the next line after an initial response seemed to induce *ESR1* Y537N (6.26%) and reduce D538G (8.73%), which were detected at disease progression (**Figure 4B**).

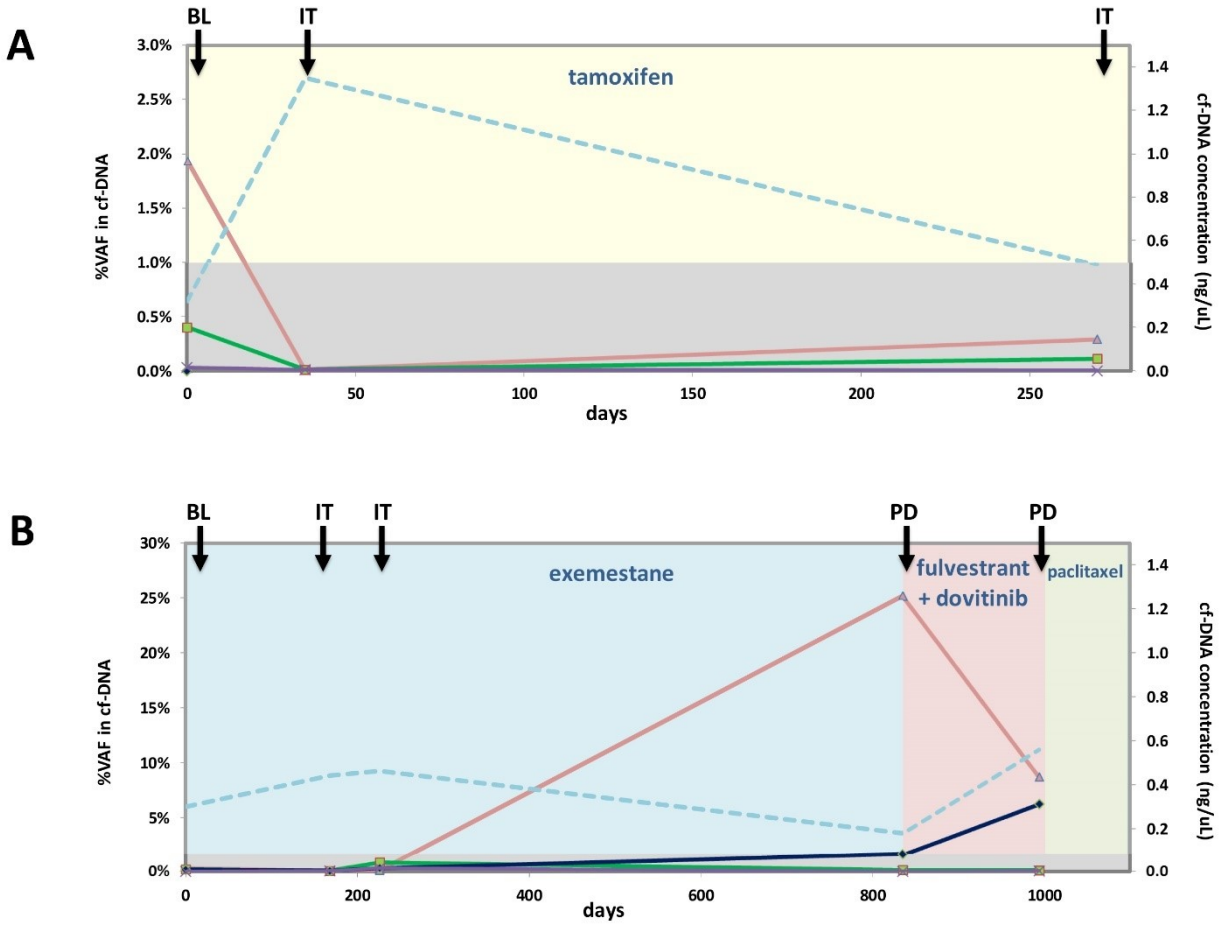


Figure 4: Longitudinal monitoring changes in cell-free DNA (cfDNA) concentration and *ESR1* mutational load.

Panel **A** and **B**: Longitudinal monitoring of cfDNA with respect to total concentration (ng/uL) and variant allele frequency (VAF, %) of the four m*ESR1* for two MBC patients (**A**: Patient 60, **B**; Patient 25) monitored in time by dPCR. Time is expressed as days from diagnosis of metastatic disease to plasma sampling. Details of therapies are indicated by colored shading. The dashed light blue line indicates the total concentration of cfDNA (ng/ µL). The colored lines indicate the different m*ESR1*: pink; D538G, green; Y537C, dark blue; Y537N and purple, Y537S. The grey area indicates the m*ESR1* cutoff at 1% VAF.
BL, baseline; IT, on-treatment; PD, progressive disease.

Discussion

cfDNA offers several important advantages for real-time monitoring of a tumor response to therapy, and as a result, an increasing number of studies accessing cfDNA through a minimally invasive blood draw has been reported in the last few years [17, 34, 35]. However, published studies often give remarkable importance to clinical data without providing details about the conditions under which the clinical samples were analyzed. The main purpose of this study was to set up a reliable, easy-to-implement workflow for uniplex and multiplex mutation analyses in cfDNA derived from minute amounts of plasma.

In general, three phases are relevant to establish an assay: the pre-analytical, analytical and post-analytical phase. It is important to realize that all these three phases are prone to errors [36-38]. Although several studies reported a higher time-dependent increase of degenerated lysed white blood cells in EDTA tubes during the pre-analytical phase [39, 40], our group previously showed that this increase is not significant up to 24 hours from blood draw [41]. Therefore, to reduce the risk of contamination with DNA derived from white blood cells, the input material we restricted to plasma collected within 24 hours from blood draw. Regarding the analytical phase, several challenges relate to the robust and reproducible analysis of m*ESR1* by dPCR in small amounts of cfDNA derived from cancer patients. To date, several studies have reported that dPCR is a highly sensitive method to detect *ESR1* mutations in cfDNA [3, 5-7, 12, 18-20]. However, cfDNA concentrations and/or total input may affect the mutation detection rate at a low VAF and the possibility to perform multiple analyses on the same sample [42, 43]. It is known that cfDNA is highly fragmented (approximately 150 bp) and usually constitutes only a small percentage of total cfDNA (often < 1%) [44, 45]. As a consequence, optimal recovery efficiency and quality of cfDNA is usually only obtained from a sufficient amount of blood plasma (>1 mL) [46-48]. For retrospective cohorts, the volume of an available sample is often limited (in this study, sometimes only 200 µL plasma samples were available), yielding a relatively lower amount of cfDNA (in this study, ranging between 2 to 200 ng). The generation of reliable data for multiple variants requires an unbiased pre-amplification step. Recent studies have tested the amplification of small amounts of DNA before dPCR using whole genome amplification [49, 50]. However, the application notes of the manufacturer and our own data suggested that methods, such as REPLI-g, are not suitable for cfDNA because of their fragmented nature. For this reason, a target-specific pre-amplification step we used to increase the DNA quantity before dPCR. The experiments with different starting volumes of plasma from the same HBD, with or without pre-

amplification, clearly confirmed that an input of 200 µL plasma was sufficient for our analyses (**Table 3**). Nevertheless, despite the claim of the manufacturer (Thermo Fisher) that TaqMan PreAmp Master Mix pre-amplifies small amounts of DNA without introducing amplification bias to the sample, the possibility that the pre-amplification method used may have introduced a bias cannot be ruled out. Data have been reported regarding errors introduced by Taq Polymerase (up to 1-20 x 10^{-5} errors/bp/duplication)[51] that may be the source of false-positive wells in the dPCR chip [50]. This might explain the presence of some mutant copies detected in cfDNA from pre-amplified healthy blood donor samples, which were not detected in the matched not pre-amplified samples. Therefore, to prevent false-positives, a stringent cutoff value at 1% was used for all of the four m*ESR1,* which was well above the average VAF + 2.58 x SD measured in 18 individual healthy blood donors.

Besides optimizing the target-specific pre-amplification step to enable mutation-specific analyses in small volumes of plasma, a custom multiplex m*ESR1* screening dPCR approach was also introduced in the workflow. The possibility to detect multiple mutations in parallel has already been demonstrated with droplet digital PCR in combination with, for example, the *KRAS* Screening Multiplex Kit (Biorad, Veenendaal, The Netherlands), which is able to screen for seven *KRAS* mutations [52-54]. However, no such kits are available yet for measuring m*ESR1*. For this study, an *ESR1* multiplex dPCR screening assay for detecting and quantifying the four most common mutations in the ligand binding domain of the *ESR1* gene was, therefore, successfully developed and validated.

Finally, the outcome of our m*ESR1* workflow was compared with results obtained by targeted NGS covering hotspot mutations in 10 genes. Although the calculated variant allele frequencies were slightly different, four of the five *ESR1* mutations detected by dPCR were also identified by targeted NGS. One *ESR1* D538G mutation with a VAF of 1.9% was, probably because of the suboptimal amount of DNA available for NGS, only identified by our m*ESR1* dPCR workflow. The advantage of this m*ESR1* dPCR workflow is that far less starting material is required (down to 0.1 ng cfDNA versus at least 5 ng cfDNA for the targeted NGS).

Next, the optimized workflow was used to investigate whether a correlation between the presence of m*ESR1* and treatment existed within two retrospectively obtained patient cohorts that received chemotherapy or endocrine therapy for MBC. Although *ESR1* mutations are rarely detected in primary breast cancer [10, 55], they are known to be enriched in metastatic lesions of patients receiving endocrine therapy [3, 6]. These data are in line with these studies showing an enrichment of m*ESR1* in MBC patients treated with endocrine therapy for MBC. Althought, to

date, no direct evidence exists for the presence of these mutations in patients receiving chemotherapy, m*ESR1* were detected in three different patients who received chemotherapy and who had never received endocrine therapy, including one patient with an ER-positive primary tumor which expressed 8.2% VAF for *ESR1* Y537S in the cfDNA collected at progressive disease. Perhaps unexpected, the primary tumor of only two of these three chemotherapy treated patients with mutated m*ERS1* copies were ER-positive according to the pathology reports. Besides tumor heterogeneity and sampling bias, this finding might be explained by intrinsic tumor genetic instability that allows cancer cells to develop mutations under treatment pressure or by the presence of the mutations already in the primary tumor as previously reported by Wang et al. [5]. Furthermore, and as we already reported before [8], althought most of the m*ESR1*-positive patients who progressed on endocrine therapy received aromatase inhibitor, fulvestrant, or a combination of the two, four of the patients with mutated *ESR1* copies at progressive disease received tamoxifen single-agent therapy. These data, thus, indicate that the emergence of m*ESR1* copies is not necessarily restricted to one specific type of therapy. Nevertheless, knowing the m*ESR1* status may help patients by offering m*ESR1*-positive patients a different line of endocrine therapy to lengthen the progression free survival time. Whether this should be an ER-degrading compound in combination with, for example, a cyclin-dependent kinase 4/6 inhibitor, in line with the PALOMA-3 study [12], and if this results in a decrease in m*ESR1* remains to be seen in confirmatory studies.

Conclusions

In conclusion, we established a dPCR workflow to assess m*ESR1* in a limited amount of cfDNA, which was compared with targeted NGS. This workflow was successfully used to investigate the *ESR1* mutational status in a retrospectively collected cohort of MBC patients, treated with either endocrine or chemotherapy, and showed a clear enrichment in m*ESR1* in samples of patients who progressed on endocrine treatment.

References

1. Riggins, R.B., et al., *Pathways to tamoxifen resistance.* Cancer Lett, 2007. **256**(1): p. 1-24.
2. Robinson, D.R., et al., *Activating ESR1 mutations in hormone-resistant metastatic breast cancer.* Nat Genet, 2013. **45**(12): p. 1446-51.
3. Guttery, D.S., et al., *Noninvasive detection of activating estrogen receptor 1 (ESR1) mutations in estrogen receptor-positive metastatic breast cancer.* Clin Chem, 2015. **61**(7): p. 974-82.
4. Angus, L., et al., *ESR1 mutations: Moving towards guiding treatment decision-making in metastatic breast cancer patients.* Cancer Treat Rev, 2017. **52**: p. 33-40.
5. Wang, P., et al., *Sensitive Detection of Mono- and Polyclonal ESR1 Mutations in Primary Tumors, Metastatic Lesions, and Cell-Free DNA of Breast Cancer Patients.* Clin Cancer Res, 2016. **22**(5): p. 1130-7.
6. Schiavon, G., et al., *Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer.* Sci Transl Med, 2015. **7**(313): p. 313ra182.
7. Chandarlapaty, S., et al., *Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial.* JAMA Oncol, 2016. **2**(10): p. 1310-1315.
8. Beije, N., et al., *Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients.* Mol Oncol, 2018. **12**(1): p. 48-57.
9. Cancer Genome Atlas, N., *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
10. Toy, W., et al., *ESR1 ligand-binding domain mutations in hormone-resistant breast cancer.* Nat Genet, 2013. **45**(12): p. 1439-45.
11. Takeshita, T., et al., *Clinical significance of plasma cell-free DNA mutations in PIK3CA, AKT1, and ESR1 gene according to treatment lines in ER-positive breast cancer.* Molecular Cancer, 2018. **17**.
12. Fribbens, C., et al., *Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer.* J Clin Oncol, 2016. **34**(25): p. 2961-8.
13. O'Leary, B., et al., *Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer.* Nat Commun, 2018. **9**(1): p. 896.

14. Merenbakh-Lamin, K., et al., *D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer.* Cancer Res, 2013. **73**(23): p. 6856-64.
15. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. **366**(10): p. 883-892.
16. Yates, L.R., et al., *Subclonal diversification of primary breast cancer revealed by multiregion sequencing.* Nat Med, 2015. **21**(7): p. 751-9.
17. Diaz, L.A., Jr. and A. Bardelli, *Liquid biopsies: genotyping circulating tumor DNA.* J Clin Oncol, 2014. **32**(6): p. 579-86.
18. Takeshita, T., et al., *Droplet digital polymerase chain reaction assay for screening of ESR1 mutations in 325 breast cancer specimens.* Transl Res, 2015. **166**(6): p. 540-553 e2.
19. Chu, D., et al., *ESR1 Mutations in Circulating Plasma Tumor DNA from Metastatic Breast Cancer Patients.* Clin Cancer Res, 2016. **22**(4): p. 993-9.
20. Takeshita, T., et al., *Clinical significance of monitoring ESR1 mutations in circulating cell-free DNA in estrogen receptor positive breast cancer patients.* Oncotarget, 2016. **7**(22): p. 32504-18.
21. Hrebien, S., et al., *Reproducibility of Digital PCR Assays for Circulating Tumor DNA Analysis in Advanced Breast Cancer.* PLoS One, 2016. **11**(10): p. e0165023.
22. Gelsomino, L., et al., *ESR1 mutations affect anti-proliferative responses to tamoxifen through enhanced cross-talk with IGF signaling.* Breast Cancer Res Treat, 2016. **157**(2): p. 253-265.
23. Takeshita, T., et al., *Comparison of ESR1 Mutations in Tumor Tissue and Matched Plasma Samples from Metastatic Breast Cancer Patients.* Transl Oncol, 2017. **10**(5): p. 766-771.
24. Yanagawa, T., et al., *Detection of ESR1 mutations in plasma and tumors from metastatic breast cancer patients using next-generation sequencing.* Breast Cancer Res Treat, 2017. **163**(2): p. 231-240.
25. Masunaga, N., et al., *Highly sensitive detection of ESR1 mutations in cell-free DNA from patients with metastatic breast cancer using molecular barcode sequencing.* Breast Cancer Res Treat, 2018. **167**(1): p. 49-58.
26. Sieuwerts, A.M., et al., *mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients.* Clin Cancer Res, 2011. **17**(11): p. 3600-18.

27. Mostert, B., et al., *Gene expression profiles in circulating tumor cells to predict prognosis in metastatic breast cancer patients.* Ann Oncol, 2015. **26**(3): p. 510-6.
28. Reijm, E.A., et al., *An 8-gene mRNA expression profile in circulating tumor cells predicts response to aromatase inhibitors in metastatic breast cancer patients.* BMC Cancer, 2016. **16**: p. 123.
29. Sieuwerts, A.M., et al., *Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR.* Breast Cancer Res Treat, 2009. **118**(3): p. 455-68.
30. Shin, S., et al., *Validation and optimization of the Ion Torrent S5 XL sequencer and Oncomine workflow for BRCA1 and BRCA2 genetic testing.* Oncotarget, 2017. **8**(21): p. 34858-34866.
31. Li, X. and C. Zhou, *Comparison of cross-platform technologies for EGFR T790M testing in patients with non-small cell lung cancer.* Oncotarget, 2017. **8**(59): p. 100801-100818.
32. Jansen, M.P., et al., *Cell-free DNA mutations as biomarkers in breast cancer patients receiving tamoxifen.* Oncotarget, 2016. **7**(28): p. 43412-43418.
33. Weerts, M.J.A., et al., *Somatic Tumor Mutations Detected by Targeted Next Generation Sequencing in Minute Amounts of Serum-Derived Cell-Free DNA.* Sci Rep, 2017. **7**(1): p. 2136.
34. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics.* Nat Med, 2008. **14**(9): p. 985-90.
35. Crowley, E., et al., *Liquid biopsy: monitoring cancer-genetics in the blood.* Nat Rev Clin Oncol, 2013. **10**(8): p. 472-84.
36. Plebani, M., *The detection and prevention of errors in laboratory medicine.* Ann Clin Biochem, 2010. **47**(Pt 2): p. 101-10.
37. Plebani, M., M. Laposata, and G.D. Lundberg, *The brain-to-brain loop concept for laboratory testing 40 years after its introduction.* Am J Clin Pathol, 2011. **136**(6): p. 829-33.
38. Hawkins, R., *Managing the pre- and post-analytical phases of the total testing process.* Ann Lab Med, 2012. **32**(1): p. 5-16.
39. Sherwood, J.L., et al., *Optimised Pre-Analytical Methods Improve KRAS Mutation Detection in Circulating Tumour DNA (ctDNA) from Patients with Non-Small Cell Lung Cancer (NSCLC).* PLoS One, 2016. **11**(2): p. e0150197.

40. Parpart-Li, S., et al., *The Effect of Preservative and Temperature on the Analysis of Circulating Tumor DNA.* Clin Cancer Res, 2017. **23**(10): p. 2471-2477.
41. van Dessel, L.F., et al., *Application of circulating tumor DNA in prospective clinical oncology trials - standardization of preanalytical conditions.* Mol Oncol, 2017. **11**(3): p. 295-304.
42. Vogelstein, B. and K.W. Kinzler, *Digital PCR.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 9236-41.
43. Hudecova, I., *Digital PCR analysis of circulating nucleic acids.* Clin Biochem, 2015. **48**(15): p. 948-56.
44. Diehl, F., et al., *Detection and quantification of mutations in the plasma of patients with colorectal tumors.* Proc Natl Acad Sci U S A, 2005. **102**(45): p. 16368-73.
45. Holdhoff, M., et al., *Analysis of circulating tumor DNA to confirm somatic KRAS mutations.* J Natl Cancer Inst, 2009. **101**(18): p. 1284-5.
46. Fleischhacker, M., et al., *Methods for isolation of cell-free plasma DNA strongly affect DNA yield.* Clin Chim Acta, 2011. **412**(23-24): p. 2085-8.
47. Page, K., et al., *Influence of plasma processing on recovery and analysis of circulating nucleic acids.* PLoS One, 2013. **8**(10): p. e77963.
48. Devonshire, A.S., et al., *Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification.* Anal Bioanal Chem, 2014. **406**(26): p. 6499-512.
49. Reid, A.L., et al., *Detection of BRAF-V600E and V600K in melanoma circulating tumour cells by droplet digital PCR.* Clin Biochem, 2015. **48**(15): p. 999-1002.
50. Jackson, J.B., et al., *Multiplex Preamplification of Serum DNA to Facilitate Reliable Detection of Extremely Rare Cancer Mutations in Circulating DNA by Digital PCR.* J Mol Diagn, 2016. **18**(2): p. 235-43.
51. McInerney, P., P. Adams, and M.Z. Hadi, *Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase.* Mol Biol Int, 2014. **2014**: p. 287430.
52. Pender, A., et al., *Efficient Genotyping of KRAS Mutant Non-Small Cell Lung Cancer Using a Multiplexed Droplet Digital PCR Approach.* PLoS One, 2015. **10**(9): p. e0139074.
53. Denis, J.A., et al., *Droplet digital PCR of circulating tumor cells from colorectal cancer patients can predict KRAS mutations before surgery.* Mol Oncol, 2016. **10**(8): p. 1221-31.
54. Sorber, L., et al., *A Comparison of Cell-Free DNA Isolation Kits: Isolation and Quantification of Cell-Free DNA in Plasma.* J Mol Diagn, 2017. **19**(1): p. 162-168.

55. Jeselsohn, R., et al., *Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer.* Clin Cancer Res, 2014. **20**(7): p. 1757-1767.

Table S1: Details clinical cohorts

Clinical parameter	Description	First line chemotherapy 248 study (n=40)	First line endocrine therapy 248 study (n=81)	>first line endocrine therapy 405 study (n=35)
Number of patients included	Baseline	16	51	0
	On-treatment	22	18	2
	Progressive disease	2	12	33
Age at plasma sampling, median years (range)	Baseline	56 (30-86)	67 (37-89)	
	On-treatment	58 (35-76)	58 (37-91)	67 (48-86)
	Progressive disease	43 (34-51)	64 (37-82)	68 (35-88)
Time between meta and plasma sampling, months (range)	Baseline	1 (0-5)	0 (0-3)	
	On-treatment	1 (0-43)	0 (0-3)	16 (4-28)
	Progressive disease	3 (0-5)	0 (0-1)	25 (1-183)
Therapy lines prior/during plasma sampling	None	9	32	0
	AI only	0	22	17
	Tamoxifen only	1	10	7
	AI + tamoxifen	0	1	6
	AI + targetted	0	1	0
	Fulvestrant + targetted	0	2	1
	Endocrine + chemo/targetted	2	12	1
	Anthracyclines only	8	0	2
	Taxanes only	13	0	0
	Combined or other chemotherapy	7	1	1

Table S2: Detailed protocols used for the pre-amplification and digital PCR

Pre-amplification	**UNIPLEX ASSAY**	**MULTIPLEX ASSAY**	
	µL	µL	
PreAmp Mastermix*	2.13	2.13	
PreAmp assay 1:100†	1.07	1.07	
DNA (0.1-2 ng)	2.00	2.00	
final volume	5.20	5.20	
10 min 95°C			
30 sec 95°C + 4 min 60°C (for 15 cycle)			
0°C			
dPCR	**UNIPLEX ASSAY**	***ESR1* MULTIPLEX SCREENING ASSAY**	
	µL	µL	
dPCR Mastermix [2x]‡	8.70	8.70	
Taqman assay 40x	0.85	0.85	D538G
		0.85	Y537S
		0.43	Y537C+Y537N§
DNA (20-40 ng)	7.85	7.85	
final volume	17.40	18.68	
load 14.5 uL in the dPCR chip			
10 min 96°C			
30 sec 98°C + 2 min 55°C	40x		
10°C			

**Catalog Number ThermoFisher: 4488593*

† Dilute Taqman/primer assay 100x	
Number of assays, eg	10
µL of tach Taqman assay [40x]/primer [100 µM]	1
µL LoTe	90
‡ QuantStudio 3D Digital PCR Master Mix v2,	
Catalog number Thermo Fisher: A26358	
§ Y537C+Y537N	
18 µL 100 uM F	
18 µL 100 uM R	
5 µL FAM probe Y537C [40x]	
5 µL FAM probe Y537N [40x]	
5 µL VIC probe Y537C/Y537N [40x]	

Table S3: Calculating input pre-amplified copies for dPCR

Cq WT *ESR1*_Y537S (qPCR)	Dilution and µL pre-amplified sample in the dPCR			
	1:10	1:100	1:1,000	1:10,000
28	**8**	80	800	8000
27	**4**	40	400	4000
26	**2**	20	200	2000
25	1	**10**	100	1000
24	0.5	**5**	50	500
23	0.25	**2.5**	25	250
22	0.125	**1.25**	12.5	125
21	0.0625	0.625	**6.25**	62.5
20	0.0313	0.313	**3.13**	31.3
19	0.0156	0.156	**1.56**	15.6
18	0.00781	0.0781	0.781	**7.8**
17	0.00391	0.0391	0.391	**3.9**
16	0.00195	0.0195	0.195	**2.0**
15	0.00098	0.0098	0.098	**1.0**

WT, wild-type; Cq, PCR cycle at the pre-specified threshold for quantification.

The optimal input of µL pre-amplified material at the specified dilution is in bold.

Table S4: Detailed NGS data

Sample		Oncomine NGS							
ID patients*	Date of plasma sampling	cfDNA input (ng)	Bases	≥ Q20	Reads	Mean Read Length (bp)	Median read coverage	Median Mol Cov	Targets >0.8MM Cov in %
12	25-oct-13	**< 1**	44937126	42757067	476296	94	16020	362	92
12	29-nov-13	6.0	114930486	109091568	1243039	92	41047	699	76
16	18-jul-12	9.3	52112485	48561077	849746	61	15221	772	88
16	12-nov-14	**3.4**	75813387	71357913	977016	78	26271	402	92
25	12-oct-10	**3.9**	2253725	2149948	33719	67	640	**39**	65
25	29-mar-11	5.3	35298423	33381340	1109108	32	1668	127	88
25	26-may-11	5.5	57370443	53464800	985604	58	15723	209	80
25	24-jan-13	**2.2**	16974025	16132767	198980	85	6434	145	65
42	8-sep-11	**0.9**	7958695	7538220	127079	63	2230	**31**	69
42	23-jan-12	6.1	154635655	142860216	1991914	78	49126	935	84
51	21-aug-12	**4.2**	135436	127960	1505	90	32	**4**	53
51	28-feb-14	5.9	61476372	57268432	1114859	55	16387	220	88
60	8-dec-11	**3.2**	7431466	7027190	111378	67	2276	**46**	69
60	12-jan-12	6.1	37194042	34913354	900268	41	5695	**70**	76
60	3-sep-12	5.9	60864293	57374	1119410	54	15108	175	84
77	6-nov-12	**4.3**	58164262	55306217	672841	86	20101	253	69
77	20-feb-14	6.0	150118054	138755257	1851389	81	51101	888	92
HBDs*									
1	22-dec-16	7.8	41654268	78443892	1228238	69	25921	988	65
2	22-dec-16	6.4	108795173	101384392	1314819	83	39757	641	88
3	22-dec-16	6.7	61309124	56780452	979627	63	19358	626	88
4	22-dec-16	9.9	87591282	80963592	1288132	68	28412	1527	88
5	22-dec-16	5.6	21252054	19446337	617358	34	1817	256	76
6	22-dec-16	9.1	99044135	91496171	1517058	65	28552	927	88
7	22-dec-16	10.4	88266607	81686095	1353171	65	26804	14463	88
8	22-dec-16	**4.9**	32946240	31228556	660280	50	7345	450	76
9	22-dec-16	6.1	52142614	49510294	857020	61	14406	581	88
10	22-dec-16	6.9	78474033	74072688	1185024	66	24688	771	76

Input less than 5 ng is noted in bold.

HBDs, Healthy blood donors; Q20, Quality Scores; Mol Cov, Molecular Coverage; MM Cov, Median Molecular Coverage; * anonymized patients/HBDs number

Table S5: Detailed optimization data

Experiment description

Pre-amp with 0.5 µL cfDNA input and UNIPLEX Pre-amplification protocol													
Sample ID	**1.µL PREAMP**	**1.Qubit Conc Eluate (ng/µL)**	**1.PRE-AMP PRIMERS**	**1.Cq value**	**1. Average input after pre-amplification (ng)**	**1.*ESR1* D538G AVG MUTATED (%)**	**1.*ESR1* Y537C AVG MUTATED (%)**	**1.*ESR1* Y537N AVG MUTATED (%)**	**1.*ESR1* Y537S AVG MUTATED (%)**	**1.*ESR1* D538G SD MUTATED (%)**	**1.*ESR1* Y537C SD MUTATED (%)**	**1.*ESR1* Y537N SD MUTATED (%)**	**1.*ESR1* Y537S SD MUTATED (%)**
9	0,5	**0,40**	UNIPLEX	**22,17**	**10,6**	**27,11**	**12,08**	0,13	0,00	0,21	0,08	0,00	0,00
15	0,5	**0,69**	UNIPLEX	**19,39**	**16,8**	**10,18**	0,10	0,32	0,06	0,09	0,02	0,02	0,02
16	0,5	**0,84**	UNIPLEX	**19,01**	**20,9**	0,24	0,12	**6,30**	0,02	0,01	0,00	0,03	0,02
25	0,5	**0,56**	UNIPLEX	**20,13**	**6,5**	**15,98**	0,12	0,23	0,04	0,24	0,00	0,08	0,05
10	0,5	**0,54**	UNIPLEX	**20,26**	**11,6**	**2,86**	0,25	0,28	0,02	0,17	0,01	0,03	0,03
11	0,5	**0,42**	UNIPLEX	**23,51**	**14,5**	**11,27**	0,45	0,24	0,02	0,14	0,12	0,04	0,01
13	0,5	**0,54**	UNIPLEX	**20,62**	**11,2**	**5,14**	0,13	0,00	0,06	0,17	0,08	0,00	0,01
24	0,5	**0,79**	UNIPLEX	**20,19**	**12,7**	**20,55**	0,16	0,15	0,06	0,10	0,01	0,00	0,00
3	0,5	**0,46**	UNIPLEX	**22,16**	**9,0**	0,22	0,00	0,14	0,14	0,02	0,00	0,04	0,03
5	0,5	**0,47**	UNIPLEX	**19,19**	**11,6**	0,40	0,41	0,20	0,02	0,02	0,03	0,06	0,02
1	0,5	**0,44**	UNIPLEX	**21,55**	**7,8**	0,18	0,07	0,03	0,02	0,00	0,10	0,04	0,03
2	0,5	**7,79**	UNIPLEX	**16,15**	**74,1**	0,12	0,14	0,10	0,04	0,02	0,02	0,01	0,01
4	0,5	**0,50**	UNIPLEX	**19,63**	**9,5**	0,12	0,28	0,03	0,04	0,05	0,00	0,05	0,06
6	0,5	**0,50**	UNIPLEX	**20,52**	**10,6**	0,13	0,09	0,36	0,04	0,06	0,09	0,02	0,06
7	0,5	**1,86**	UNIPLEX	**18,87**	**14,8**	0,31	0,20	0,06	0,04	0,01	0,05	0,00	0,02
8	0,5	**0,93**	UNIPLEX	**19,82**	**14,3**	0,45	0,18	0,16	0,05	0,09	0,02	0,02	0,03
12	0,5	**2,24**	UNIPLEX	**16,45**	**22,1**	0,18	0,23	0,11	0,05	0,00	0,03	0,02	0,01
14	0,5	**0,35**	UNIPLEX	**20,30**	**5,2**	0,73	0,23	0,07	0,10	0,15	0,01	0,09	0,00
17	0,5	**0,38**	UNIPLEX	**22,02**	**8,2**	0,54	0,08	0,44	0,13	0,11	0,12	0,00	0,04
18	0,5	**0,64**	UNIPLEX	**20,78**	**10,5**	0,21	0,19	0,06	0,00	0,06	0,00	0,03	0,00
19	0,5	**0,78**	UNIPLEX	**19,89**	**15,6**	0,27	0,36	0,14	0,06	0,00	0,05	0,04	0,01
20	0,5	**0,70**	UNIPLEX	**20,55**	**10,0**	0,12	0,34	0,10	0,03	0,04	0,03	0,00	0,01
21	0,5	**0,70**	UNIPLEX	**20,01**	**8,2**	0,28	0,18	0,03	0,09	0,04	0,04	0,05	0,12
22	0,5	**0,37**	UNIPLEX	**23,71**	**9,6**	0,06	0,10	0,52	0,11	0,00	0,04	0,08	0,01
23	0,5	**0,46**	UNIPLEX	**21,25**	**13,6**	0,13	0,04	0,18	0,05	0,04	0,00	0,03	0,02
26	0,5	**0,35**	UNIPLEX	**20,55**	**4,6**	0,36	0,22	0,05	0,47	0,16	0,06	0,07	0,10
27	0,5	**0,82**	UNIPLEX	**19,56**	**9,4**	0,28	0,31	0,21		0,03	0,05	0,03	
28	0,5	**0,62**	UNIPLEX	**20,45**	**32,6**	0,24	0,10	0,26		0,03	0,03	0,00	
29	0,5	**1,66**	UNIPLEX	**16,58**	**23,6**	0,37	0,12	0,10		0,01	0,00	0,01	
30	0,5	**0,56**	UNIPLEX	**22,41**	**25,9**	0,08	0,09	0,35			0,00	0,04	
31	0,5	**0,90**	UNIPLEX	**18,96**	**16,9**		0,08	0,08			0,00	0,00	

Samples positive for one or more of the m*ESR1* mutations are in bold.

Pre-amp with 2 µL cfDNA input and MULTIPLEX Pre-amplification protocol															
Sample ID	**2.Input Pre-amp (µL)**	**2.Qubit Conc Eluate (ng/µL)**	**2.PRE-AMP PRIMERS**	**2.Cq value**	**2. Average input after pre-amplification (ng)**	**2.ESR1 D538G AVG MUTATED (%)**	**2.ESR1 Y537C AVG MUTATED (%)**	**2.ESR1 Y537N AVG MUTATED (%)**	**2.ESR1 Y537S AVG MUTATED (%)**	**2.ESR1 G_ C_N_S AVG MUTATED (%)**	**2.ESR1 D538G SD MUTATED (%)**	**2.ESR1 Y537C SD MUTATED (%)**	**2.ESR1 Y537N SD MUTATED (%)**	**2.ESR1 Y537S SD MUTATED (%)**	**2.ESR1 G_C_N_S SD MUTATED (%)**
9	2,0	**0,40**	MULTI	**20,76**	**10,6**	**14,63**	**3,85**	0,04	0,16	**18,28**	0,35	0,10	0,00	0,06	0,23
15	2,0	**0,69**	MULTI	**18,07**	**10,6**	**17,80**	0,27	0,17	0,04	**18,90**	0,13	0,03	0,00	0,00	0,03
16	2,0	**0,84**	MULTI	**17,77**	**14,1**	0,18	0,29	**6,39**	0,00	**5,39**	0,07	0,03	0,02	0,00	0,21
25	2,0	**0,56**	MULTI	**19,80**	**11,5**	**8,73**	0,10	**6,26**	0,02		0,04	0,03	0,03	0,03	
10	2,0	**0,54**	MULTI	**20,12**	**9,9**	**5,97**	0,39	0,68	0,05		0,04	0,03	0,03	0,00	
11	2,0	**0,42**	MULTI	**21,06**	**9,5**	**11,04**	0,46	0,73	0,02		0,15	0,07	0,00	0,03	
13	2,0	**0,54**	MULTI	**18,74**	**6,8**	**5,08**	0,17	0,43	0,00		0,01	0,05	0,15	0,00	
24	2,0	**0,79**	MULTI	**19,14**	**13,7**	**20,50**	0,04	0,08	0,02		0,51	0,02	0,02	0,02	
3	2,0	**0,46**	MULTI	**19,56**	**11,8**	0,11	**2,45**	0,15	0,23	**2,96**	0,01	0,07	0,11	0,11	0,07
5	2,0	**0,47**	MULTI	**19,49**	**11,1**	0,15	0,41	0,21	0,00	**2,04**	0,01	0,02	0,01	0,00	0,35
1	2,0	**0,44**	MULTI	**20,81**	**11,4**	0,00	0,20	0,06	0,00		0,00	0,01	0,03	0,00	
2	2,0	**7,79**	MULTI	**15,13**	**21,0**	0,29	0,17	0,15	0,00		0,10	0,01	0,02	0,00	
4	2,0	**0,50**	MULTI	**18,37**	**11,5**	0,07	0,12	0,06	0,20		0,00	0,00	0,03	0,13	
6	2,0	**0,50**	MULTI	**19,68**	**13,2**	0,21	0,41	0,20	0,07		0,02	0,07	0,02	0,09	
7	2,0	**1,86**	MULTI	**17,26**	**28,2**	0,21	0,07	0,10	0,02		0,00	0,00	0,01	0,00	
8	2,0	**0,93**	MULTI	**18,14**	**17,7**	0,25	0,12	0,12	0,00		0,02	0,03	0,00	0,00	
12	2,0	**2,24**	MULTI	**16,03**	**11,1**	0,47	0,14	0,06	0,00		0,05	0,03	0,03	0,00	
14	2,0	**0,35**	MULTI	**19,63**	**11,7**	0,31	0,13	0,30	0,49		0,01	0,03	0,09	0,20	
17	2,0	**0,38**	MULTI	**21,48**	**8,7**	0,24	0,10	0,50	0,61		0,07	0,00	0,03	0,11	
18	2,0	**0,64**	MULTI	**18,93**	**8,6**	0,38	0,63	0,33	0,00		0,18	0,06	0,16	0,00	
19	2,0	**0,78**	MULTI	**19,41**	**16,1**	0,21	0,17	0,08	0,00		0,02	0,00	0,00	0,00	
20	2,0	**0,70**	MULTI	**18,95**	**9,2**	0,54	0,20	0,12	0,00		0,01	0,04	0,03	0,00	
21	2,0	**0,70**	MULTI	**18,94**	**12,3**	0,32	0,17	0,02	0,02		0,02	0,05	0,03	0,02	
22	2,0	**0,37**	MULTI	**21,90**	**9,2**	0,85	0,09	0,07	0,00		0,07	0,00	0,04	0,00	
23	2,0	**0,46**	MULTI	**20,20**	**8,8**	0,02	0,19	0,00	0,00		0,03	0,04	0,00	0,00	
26	2,0	**0,35**	MULTI	**20,63**	**10,9**	0,57	0,12	0,11	0,00		0,03	0,00	0,03	0,00	
27	2,0	**0,82**	MULTI	**18,45**	**12,8**	0,25	0,12	0,14	0,04		0,01	0,02	0,03	0,00	
28	2,0	**0,62**	MULTI	**19,69**	**14,9**	0,07	0,22	0,15	0,00		0,02	0,04	0,02	0,00	
29	2,0	**1,66**	MULTI	**15,18**	**9,5**	0,27	0,32	0,34	0,03		0,00	0,00	0,06	0,04	
30	2,0	**0,56**	MULTI	**20,01**	**6,1**	0,00	0,04	0,35	0,33		0,00	0,06	0,23	0,00	
31	2,0	**0,90**	MULTI	**17,94**	**16,0**	0,21	0,08	0,13	0,01		0,00	0,00	0,00	0,02	

Pre-amp with 2 µL cfDNA input and UNIPLEX Pre-amplification protocol													
Sample ID	**1.µL PREAMP**	**1.Qubit Conc Eluate (ng/µL)**	**1.PRE-AMP PRIMERS**	**1.Cq value**	**1. Average input after pre-amplification (ng)**	**1.ESR1 D538G AVG MUTATED (%)**	**1.ESR1 Y537C AVG MUTATED (%)**	**1.ESR1 Y537N AVG MUTATED (%)**	**1.ESR1 Y537S AVG MUTATED (%)**	**1.ESR1 D538G SD MUTATED (%)**	**1.ESR1 Y537C SD MUTATED (%)**	**1.ESR1 Y537N SD MUTATED (%)**	**1.ESR1 Y537S SD MUTATED (%)**
32	2,0	**0,33**	UNIPLEX	**20,53**	**7,3**	0,32	0,24	0,10	**8,36**	0,04	0,00	0,04	0,16
33	1,0	**0,78**	UNIPLEX	**21,09**	**13,4**	0,12			**7,17**	0,02			0,06
40	2,0	**0,49**	UNIPLEX	**19,88**	**11,1**	0,13	0,08	0,02	0,00	0,05	0,00	0,03	0,00
27	2,0	**0,82**	UNIPLEX	**17,22**	**10,1**	0,28	0,31	0,21	0,16	0,03	0,05	0,03	0,05
28	2,0	**0,62**	UNIPLEX	**17,86**	**19,4**	0,16			0,06	0,03			0,00
30	2,0	**0,56**	UNIPLEX	**18,90**	**53,2**	0,15	0,09	0,35	0,26		0,00	0,04	0,12
31	2,0	**0,90**	UNIPLEX	**16,67**	**17,6**		0,08	0,08	0,07		0,00	0,00	0,00
34	2,0	**0,35**	UNIPLEX	**20,87**	**10,3**	0,13	0,32	0,00	0,12	0,01	0,03	0,00	0,03
35	2,0	**0,12**	UNIPLEX	**20,51**	**9,6**	0,11	0,12	0,00	0,09	0,00	0,03	0,00	0,00
36	2,0	**1,83**	UNIPLEX	**16,22**	**15,7**	0,25	0,07	0,00	0,03	0,02	0,02	0,00	0,00
37	2,0	**0,51**	UNIPLEX	**19,19**	**7,7**	0,43	0,48	0,13	0,00	0,03	0,00	0,00	0,00
38	2,0	**0,46**	UNIPLEX	**20,61**	**12,5**	0,17	0,16	0,02	0,02	0,00	0,02	0,03	0,03
39	2,0	**0,71**	UNIPLEX	**18,28**	**11,2**	0,37	0,43	0,02	0,00	0,07	0,03	0,03	0,00
41	2,0	**0,40**	UNIPLEX	**20,66**	**7,6**		0,72				0,04		
42	2,0	**0,51**	UNIPLEX	**20,77**	**11,5**		0,14				0,08		
43	2,0	**0,30**	UNIPLEX	**21,04**	**13,2**			0,09				0,11	
44	2,0	**0,89**	UNIPLEX	**21,70**	**9,7**	0,28	0,89	0,05	0,00	0,00	0,03	0,07	0,00
45	2,0	**0,39**	UNIPLEX	**21,30**	**7,9**	0,34	0,12	0,00	0,00	0,00	0,00	0,00	0,00

Samples positive for one or more of the m*ESR1* mutations are in bold.

Pre-amp with 2 µL cfDNA input and MILTIPLEX Pre-amplification protocol														
Sample ID	2.Input Pre-amp (µL)	2.Qubit Conc Eluate (ng/µL)	2.PRE-AMP PRIMERS	2.Cq value	2. Average input after pre-amplification (ng)	2.ESR1 D538G AVG MUTATED (%)	2.ESR1 Y537C AVG MUTATED (%)	2.ESR1 Y537N AVG MUTATED (%)	2.ESR1 Y537S AVG MUTATED (%)	2.ESR1 G C N S AVG MUTATED (%)	2.ESR1 D538G SD MUTATED (%)	2.ESR1 Y537C SD MUTATED (%)	2.ESR1 Y537N SD MUTATED (%)	2.ESR1 Y537S SD MUTATED (%)
32	2,0	**0,33**	MULTI	**22,01**	**11,8**				**10,70**					0,13
33	2,0	**0,78**	MULTI	**18,98**	**7,2**	0,60	0,19	0,21	**2,02**		0,09	0,09	0,04	0,24
40	2,0	**0,49**	MULTI	**20,63**	**9,0**	**1,71**	0,10	0,19	0,06		0,11	0,00	0,00	0,00
27	2,0	**0,82**	MULTI	**18,45**	**12,8**	0,25	0,12	0,14	0,04		0,01	0,02	0,03	0,00
28	2,0	**0,62**	MULTI	**19,69**	**14,9**	0,07	0,22	0,15	0,00		0,02	0,04	0,02	0,00
30	2,0	**0,56**	MULTI	**20,01**	**6,1**	0,00	0,04	0,35	0,33		0,00	0,06	0,23	0,00
31	2,0	**0,90**	MULTI	**17,94**	**16,0**	0,21	0,08	0,13	0,01		0,00	0,00	0,00	0,02
34	2,0	**0,35**	MULTI	**21,58**	**11,0**	0,12	0,36	0,00	0,19		0,11	0,00	0,00	0,06
35	2,0	**0,12**	MULTI	**20,95**	**7,5**	0,27	0,16	0,13	0,00		0,02	0,04	0,04	0,00
36	2,0	**1,83**	MULTI	**15,76**	**11,7**	0,09	0,25	0,08	0,00		0,00	0,09	0,02	0,00
37	2,0	**0,51**	MULTI		**7,4**	0,74	0,17	0,00	0,00		0,09	0,00	0,00	0,00
38	2,0	**0,46**	MULTI	**19,70**	**10,8**	0,29	0,44	0,11	0,00		0,00	0,00	0,03	0,00
39	2,0	**0,71**	MULTI	**17,88**	**7,3**	0,39	0,15	0,09	0,00		0,05	0,04	0,04	0,00
41	2,0	**0,40**	MULTI	**20,25**	**6,5**	0,10	0,28	0,00	0,07		0,05	0,00	0,00	0,00
42	2,0	**0,51**	MULTI	**20,66**	**7,9**	0,25	0,42	0,05	0,15		0,04	0,00	0,00	0,00
43	2,0	**0,30**	MULTI	**21,30**	**9,6**	0,30	0,18	0,17	0,00		0,03	0,00	0,04	0,00
44	2,0	**0,89**	MULTI	**21,20**	**7,8**	0,90	0,13	0,05	0,00		0,07	0,02	0,07	0,00
45	2,0	**0,39**	MULTI	**21,59**	**7,1**	0,35	0,48	0,00	0,00		0,00	0,01	0,00	0,00

Digital PCR multiplex screening															
Sample ID	**1.µL PREAMP**	**1.Qubit Conc Eluate (ng/µL)**	**1.PRE-AMP PRIMERS**	**1.Cq value**	**1. Average input after pre-amplification (ng)**	**1.ESR1 D538G AVG MUTATED (%)**	**1.ESR1 Y537C AVG MUTATED (%)**	**1.ESR1 Y537N AVG MUTATED (%)**	**1.ESR1 Y537S AVG MUTATED (%)**	**1.ESR1 G_ C_ N_S AVG MUTATED (%)**	**1.ESR1 D538G SD MUTATED (%)**	**1.ESR1 Y537C SD MUTATED (%)**	**1.ESR1 Y537N SD MUTATED (%)**	**1.ESR1 Y537S SD MUTATED (%)**	**1.ESR1 G_ C_ N_S SD MUTATED (%)**
9	2,0	**0,40**	MULTI	**20,76**	**10,6**	**14,63**	0,04	**3,85**	0,16	**18,28**	0,35	0,00	0,10	0,06	0,23
25	9,0	**0,56**	MULTI	**16,91**	**11,5**	**12,09**	**3,25**	0,23	0,46	**15,97**	0,07	0,14	0,09	0,03	1,42
3	2,0	**0,46**	MULTI	**19,56**	**11,8**	0,11	0,15	**2,45**	0,23	**2,96**	0,01	0,11	0,07	0,11	0,07
10	8,0	**0,54**	MULTI	**16,89**	**7,5**	**6,41**	0,90	0,09	0,00	**5,61**	0,16	0,01	0,04	0,00	0,03
13	10,0	**0,54**	MULTI	**17,06**	**11,5**	**7,07**	0,10	0,16	0,00	**6,93**	0,07	0,03	0,00	0,00	0,40
15	10,0	**0,69**	MULTI	**15,95**	**18,0**	**15,27**	0,00	0,09	0,00	**16,22**	0,05	0,00	0,02	0,00	0,02
16	2,0	**0,84**	MULTI	**17,77**	**14,1**	0,18	**6,39**	0,29	0,00	**5,39**	0,07	0,02	0,03	0,00	0,21
50	2,0	**0,63**	MULTI	**19,46**	**16,8**	0,30	0,11	0,05	23,78	**27,73**	0,04	0,00	0,00	0,05	0,27
51	2,0	**0,46**	MULTI	**22,86**	**11,2**	**2,28**	0,15	0,05	0,04	**3,11**	0,01	0,03	0,03	0,00	0,21
52	5,0	**0,42**	MULTI	**21,46**	**8,9**	0,32	**5,37**	0,16	0,07	**5,01**	0,03	0,12	0,10	0,04	1,20
53	10,0	**0,19**	MULTI	**19,01**	**6,9**	**1,69**	0,00	0,16	0,00	**2,04**	0,02	0,00	0,04	0,00	0,20
68	2,0	**0,30**	MULTI	**22,15**	**10,1**	0,08	0,05	0,22	12,08	**12,57**	0,00	0,00	0,07	0,28	0,73
70	2,0	**0,75**	MULTI	**18,83**	**11,6**	0,18	**1,22**	0,11	0,00	0,43	0,00	0,16	0,00	0,00	0,25
1	9,0	**0,44**	MULTI	**17,98**	**8,2**	0,27	0,00	0,00	0,00	0,53	0,04	0,00	0,00	0,00	0,21
17	8,0	**0,38**	MULTI	**18,65**	**7,1**	0,22	0,14	0,62	0,06	0,61	0,00	0,10	0,20	0,00	0,01
18	8,0	**0,64**	MULTI	**16,84**	**8,7**	0,33	0,08	0,25	0,03	0,58	0,00	0,04	0,00	0,04	0,13
54	2,0	**0,47**	MULTI	**20,33**	**15,9**	0,15	0,13	0,09	0,01	0,56	0,03	0,02	0,01	0,02	0,02
55	2,0	**0,81**	MULTI	**19,71**	**22,0**	0,14	0,00	0,13	0,00	0,45	0,02	0,00	0,01	0,00	0,11
56	2,0	**0,53**	MULTI	**20,91**	**15,8**	0,20	0,51	0,10	0,01	0,59	0,04	0,02	0,02	0,02	0,04
57	2,0	**0,40**	MULTI	**22,19**	**17,4**	0,02	0,05	0,15	0,00	0,36	0,00	0,00	0,05	0,00	0,06
58	2,0	**0,72**	MULTI	**18,96**	**18,9**	0,23	0,02	0,10	0,01	0,29	0,05	0,00	0,00	0,02	0,02
59	2,0	**0,44**	MULTI	**21,11**	**8,6**	0,11	0,05	0,38	0,00	0,35	0,00	0,07	0,03	0,00	0,10
60	2,0	**0,49**	MULTI	**21,59**	**13,3**	0,29	0,03	0,11	0,00	0,68	0,04	0,00	0,07	0,00	0,18
61	2,0	**0,40**	MULTI	**22,22**	**11,0**	0,15	0,18	0,02	0,00	0,52	0,05	0,04	0,03	0,00	0,03
62	2,0	**0,47**	MULTI	**20,28**	**10,7**	0,17	0,16	0,04	0,00	0,20	0,00	0,03	0,06	0,00	0,04
63	2,0	**0,42**	MULTI	**20,85**	**9,8**	0,32	0,10	0,29	0,00	**1,13**	0,01	0,00	0,00	0,00	0,04
64	2,0	**0,67**	MULTI	**20,14**	**13,4**	0,26	0,08	0,09	0,00	0,49	0,03	0,02	0,03	0,00	0,06
65	2,0	**0,45**	MULTI	**21,22**	**12,7**	0,15	0,05	0,10	0,00	0,35	0,01	0,03	0,01	0,00	0,04
66	2,0	**0,39**	MULTI	**20,99**	**10,0**	0,14	0,60	0,05	0,00	0,45	0,03	0,01	0,00	0,00	0,02
67	2,0	**0,68**	MULTI	**21,84**	**10,4**	0,13	0,15	0,00	0,02	0,27	0,02	0,03	0,00	0,03	0,05
69	2,0	**0,25**	MULTI	**20,85**	**8,4**	0,28	0,05	0,35	0,26	0,63	0,06	0,00	0,21	0,01	0,10
71	2,0	**1,01**	MULTI	**17,93**	**26,4**	0,16	0,09	0,24	0,06	0,34	0,00	0,01	0,01	0,01	0,07
72	2,0	**0,50**	MULTI	**19,01**	**10,3**	0,22	0,00	0,34	0,00	0,47	0,00	0,00	0,08	0,00	0,09
73	2,0	**0,35**	MULTI	**21,70**	**8,9**	0,24	0,05	0,17	0,07	0,57	0,00	0,00	0,03	0,03	0,03
74	10,0	**0,38**	MULTI	**20,36**	**11,4**	0,07	0,00	0,23	0,00	0,55	0,00	0,00	0,00	0,00	0,04
75	10,0	**0,10**	MULTI	**21,13**	**10,4**	0,16	0,00	0,21	0,04	0,44	0,03	0,00	0,00	0,00	0,19

Samples positive for one or more of the m*ESR1* mutations are in bold.

Reproducibility fragmented cell line															
Sample ID	**1.µL PREAMP**	**1.Qubit Conc Eluate (ng/µL)**	**1.PRE-AMP PRIMERS**	**1.Cq value**	**1. Average input after pre-amplification (ng)**	**1.ESR1 D538G AVG MUTATED (%)**	**1.ESR1 Y537C AVG MUTATED (%)**	**1.ESR1 Y537N AVG MUTATED (%)**	**1.ESR1 Y537S AVG MUTATED (%)**	**1.ESR1 G_ C_ N_S AVG MUTATED (%)**	**1.ESR1 D538G SD MUTATED (%)**	**1.ESR1 Y537C SD MUTATED (%)**	**1.ESR1 Y537N SD MUTATED (%)**	**1.ESR1 Y537S SD MUTATED (%)**	**1.ESR1 G_ C_ N_S SD MUTATED (%)**
MM435s	2,0	**0,6**	MULTIPLEX	**17,29**	**5,0**	0,15	0,00	0,04	0,00	0,78	0,07	0,00	0,06	0,00	0,15
MM435s	2,0	**0,6**	MULTIPLEX	**17,24**	**4,9**	0,15	0,00	0,00	0,00	0,72	0,08	0,00	0,00	0,00	0,09
MM435s	2,0	**0,6**	MULTIPLEX	**16,93**	**3,6**	0,58	0,49	0,59	0,05	0,67	0,00	0,00	0,15	0,07	0,14
MM435s	2,0	**0,6**	MULTIPLEX	**17,23**	**3,9**	0,18	0,33	0,45	0,00	0,65	0,09	0,00	0,18	0,00	0,16
MM435s	2,0	**0,6**	MULTIPLEX	**17,27**	**4,6**	0,15	0,00	0,00	0,00	0,38	0,07	0,00	0,00	0,00	0,07
MM435s	2,0	**0,6**	MULTIPLEX	**17,78**	**4,8**	0,19	0,26	0,31	0,00	0,37	0,00	0,00	0,07	0,00	0,12
MM435s	2,0	**0,6**	MULTIPLEX	**17,19**	**4,7**	0,25	0,27	0,35	0,00	0,29	0,07	0,00	0,12	0,00	0,03
MM435s	2,0	**0,6**	MULTIPLEX	**17,90**	**5,2**	0,24	0,00	0,00	0,08	0,22	0,00	0,00	0,00	0,12	0,01
MM435s		**11,9**	PARENTAL		**8,0**	0,00	0,45	0,47	0,00	0,20	0,00	0,11	0,10	0,00	0,10

Pre-amp with 0.5 µL cfDNA input and Speed-vac															
Sample ID	**1.µL PREAMP**	**1.Qubit Conc Eluate (ng/µL)**	**1.PRE-AMP PRIMERS**	**1.Cq value**	**1. Average input after pre-amplification (ng)**	**1.ESR1 D538G AVG MUTATED (%)**	**1.ESR1 Y537C AVG MUTATED (%)**	**1.ESR1 Y537N AVG MUTATED (%)**	**1.ESR1 Y537S AVG MUTATED (%)**	**1.ESR1 G_ C_ N_S AVG MUTATED (%)**	**1.ESR1 D538G SD MUTATED (%)**	**1.ESR1 Y537C SD MUTATED (%)**	**1.ESR1 Y537N SD MUTATED (%)**	**1.ESR1 Y537S SD MUTATED (%)**	**1.ESR1 G_C_N_S SD MUTATED (%)**
25	0,5	**0,56**	UNIPLEX	**20,13**	**6,5**	**15,98**	0,12	0,23			0,24	0,00	0,08		
10	0,5	**0,54**	UNIPLEX	**20,26**	**11,6**	**2,86**	0,25	0,28	0,02		0,17	0,01	0,03	0,03	
13	0,5	**0,54**	UNIPLEX	**20,62**	**11,2**	**5,14**	0,13	0,00	0,06		0,17	0,08	0,00	0,01	
15	0,5	**0,69**	UNIPLEX	**19,39**	**16,8**	**10,18**	0,10	0,32	0,06		0,09	0,02	0,02	0,02	
53	0,5	**0,29**	UNIPLEX	**20,94**	**6,3**	0,58			0,47		0,04			0,38	
46															
47															
48															
49															
74															
75															
1	0,5	**0,44**	UNIPLEX	**21,55**	**7,8**	0,18	0,07	0,03	0,02		0,00	0,10	0,04	0,03	
6	0,5	**0,50**	UNIPLEX	**20,52**	**10,6**	0,13		0,36	0,04		0,06		0,02	0,06	
7	0,5	**1,86**	UNIPLEX	**18,87**	**14,8**	0,31	0,20	0,06	0,04		0,01	0,05	0,00	0,02	
8	0,5	**0,93**	UNIPLEX	**19,82**	**14,3**	0,45	0,18	0,16	0,05		0,09	0,02	0,02	0,03	
12	0,5	**2,24**	UNIPLEX	**16,45**	**22,1**	0,18	0,23	0,11	0,05		0,00	0,03	0,02	0,01	
14	0,5	**0,35**	UNIPLEX	**20,30**	**5,2**	0,73	0,23	0,07	0,10		0,15	0,01	0,09	0,00	
17	0,5	**0,38**	UNIPLEX	**22,02**	**8,2**	0,54	0,08	0,44	0,13		0,11	0,12	0,00	0,04	
18	0,5	**0,64**	UNIPLEX	**20,78**	**10,5**	0,21	0,19	0,06	0,00		0,06	0,00	0,03	0,00	
19	0,5	**0,78**	UNIPLEX	**19,89**	**15,6**	0,27	0,36	0,14	0,06		0,00	0,05	0,04	0,01	
20	0,5	**0,70**	UNIPLEX	**20,55**	**10,0**	0,12	0,34	0,10	0,03		0,04	0,03	0,00	0,01	
21	0,5	**0,70**	UNIPLEX	**20,01**	**8,2**	0,28	0,18	0,03	0,09		0,04	0,04	0,05	0,12	
22	0,5	**0,37**	UNIPLEX	**23,71**	**9,6**	0,06	0,10	0,52	0,11		0,00	0,04	0,08	0,01	
23	0,5	**0,46**	UNIPLEX	**21,25**	**13,6**	0,13	0,04	0,18	0,05		0,04	0,00	0,03	0,02	
26	0,5	**0,35**	UNIPLEX	**20,55**	**4,6**	0,36	0,22	0,05	0,47		0,16	0,06	0,07	0,10	
27	0,5	**0,82**	UNIPLEX	**19,56**	**9,4**	0,28	0,31	0,21			0,03	0,05	0,03		
29	0,5	**1,66**	UNIPLEX	**16,58**	**23,6**	0,37	0,12	0,10			0,01	0,00	0,01		
51	0,5	**0,21**	MULTIPLEX	**22,33**	**7,3**	0,36				0,36	0,31				0,31

Pre-amp with 2 µL cfDNA input and Speed-vac															
Sample ID	**2.Input Pre-amp (µL)**	**2.Qubit Conc Eluate (ng/µL)**	**2.PRE-AMP PRIMERS**	**2.Cq value**	**2. Average input after pre-amplification (ng)**	**2.ESR1 D538G AVG MUTATED (%)**	**2.ESR1 Y537C AVG MUTATED (%)**	**2.ESR1 Y537N AVG MUTATED (%)**	**2.ESR1 Y537S AVG MUTATED (%)**	**2.ESR1 G_ C_ N_ S AVG MUTATED (%)**	**2.ESR1 D538G SD MUTATED (%)**	**2.ESR1 Y537C SD MUTATED (%)**	**2.ESR1 Y537N SD MUTATED (%)**	**2.ESR1 Y537S SD MUTATED (%)**	**2.ESR1 G_ C_ N_S SD MUTATED (%)**
25	2,0	**0,56**	MULTIPLEX	**19,80**	**11,5**	**8,73**	0,10	**6,26**	0,02		0,04	0,03	0,03	0,03	
10	2,0	**0,54**	MULTIPLEX	**20,12**	**9,9**	**5,97**	0,39	0,68	0,05		0,04	0,03	0,03	0,00	
13	2,0	**0,54**	MULTIPLEX	**18,74**	**6,8**	**5,08**	0,17	0,43	0,00		0,01	0,05	0,15	0,00	
15	2,0	**0,69**	MULTIPLEX	**18,07**	**9,7**	**17,80**		0,17	0,04	**18,90**	0,13		0,00	0,00	0,03
53	2,0	**0,29**	MULTIPLEX	**19,58**	**8,2**	0,18				0,76	0,03				0,13
46	2,0	**0,33**	MULTIPLEX	**22,98**	**17,5**	0,57				0,46	0,00				0,22
47	2,0	**0,34**	MULTIPLEX	**22,48**	**13,1**	0,24				0,25	0,01				0,05
48	2,0	**0,37**	MULTIPLEX	**22,38**	**13,5**					0,20					0,06
49	2,0	**0,35**	MULTIPLEX	**23,06**	**13,8**	0,27	0,36	0,19	0,00		0,00	0,08	0,00	0,00	
74	2,0	**0,38**	MULTIPLEX	**23,76**	**10,9**					0,00					0,00
75	2,0	**0,10**	UNIPLEX	**22,16**	**22,9**	0,44	0,16	0,23	0,05		0,23	0,00	0,04	0,03	
1	2,0	**0,44**	MULTIPLEX	**20,81**	**11,4**	0,00	0,20	0,06	0,00		0,00	0,01	0,03	0,00	
6	2,0	**0,50**	MULTIPLEX	**19,68**	**13,2**	0,21	0,41	0,20	0,07		0,02	0,07	0,02	0,09	
7	2,0	**1,86**	MULTIPLEX	**17,26**	**28,2**	0,21	0,07	0,10	0,02		0,00	0,00	0,01	0,00	
8	2,0	**0,93**	MULTIPLEX	**18,14**	**17,7**	0,25	0,12	0,12	0,00		0,02	0,03	0,00	0,00	
12	2,0	**2,24**	MULTIPLEX	**16,03**	**11,1**	0,47	0,14	0,06	0,00		0,05	0,03	0,03	0,00	
14	2,0	**0,35**	MULTIPLEX	**19,63**	**11,7**	0,31	0,13	0,30	0,49		0,01	0,03	0,09	0,20	
17	2,0	**0,38**	MULTIPLEX	**21,48**	**8,7**	0,24	0,10	0,50	0,61		0,07	0,00	0,03	0,11	
18	2,0	**0,64**	MULTIPLEX	**18,93**	**8,6**	0,38	0,63	0,33	0,00		0,18	0,06	0,16	0,00	
19	2,0	**0,78**	MULTIPLEX	**19,41**	**16,1**	0,21	0,17	0,08	0,00		0,02	0,00	0,00	0,00	
20	2,0	**0,70**	MULTIPLEX	**18,95**	**9,2**	0,54	0,20	0,12	0,00		0,01	0,04	0,03	0,00	
21	2,0	**0,70**	MULTIPLEX	**18,94**	**12,3**	0,32	0,17	0,02	0,02		0,02	0,05	0,03	0,02	
22	2,0	**0,37**	MULTIPLEX	**21,90**	**9,2**	0,85	0,09	0,07	0,00		0,07	0,00	0,04	0,00	
23	2,0	**0,46**	MULTIPLEX	**20,20**	**8,8**	0,02	0,19	0,00	0,00		0,03	0,04	0,00	0,00	
26	2,0	**0,35**	MULTIPLEX	**20,63**	**10,9**	0,57	0,12	0,11	0,00		0,03	0,00	0,03	0,00	
27	2,0	**0,82**	MULTIPLEX	**18,45**	**12,8**	0,25	0,12	0,14	0,04		0,01	0,02	0,03	0,00	
29	2,0	**1,66**	MULTIPLEX	**15,18**	**9,5**	0,27	0,32	0,34	0,03		0,00	0,00	0,06	0,04	
51	2,0	**0,46**	MULTIPLEX	**21,44**	**9,6**	0,23				0,73	0,03				0,00

Samples positive for one or more of the m*ESR1* mutations are in bold.

Pre-amp with different cfDNA input and Speed-vac															
Sample ID	**3.µL PREAMP**	**3.Qubit Conc Eluate (ng/µL)**	**3.PRE-AMP PRIMERS**	**3.Cq value**	**3. Average input after pre-amplification (ng)**	**3.ESR1 D538G AVG MUTATED (%)**	**3.ESR1 Y537C AVG MUTATED (%)**	**3.ESR1 Y537N AVG MUTATED (%)**	**3.ESR1 Y537S AVG MUTATED (%)**	**3.ESR1 G-C-N-S AVG MUTATED (%)**	**3.ESR1 D538G SD MUTATED (%)**	**3.ESR1 Y537C SD MUTATED (%)**	**3.ESR1 Y537N SD MUTATED (%)**	**3.ESR1 Y537S SD MUTATED (%)**	**3.ESR1 G_ C_ N_ S SD MUTATED (%)**
25	9,0	0,56	MULTIPLEX	16,91	**9,1**	**12,09**	0,23	**3,25**	0,46	**15,97**	0,07	0,09	0,14	0,03	1,42
10	8,0	0,54	MULTIPLEX	16,89	**6,2**	**6,41**	0,09	0,90	0,00	**5,61**	0,16	0,04	0,01	0,00	0,03
13	10,0	0,54	MULTIPLEX	17,06	**9,5**	**7,07**	0,16	0,10	0,00	**6,93**	0,07	0,00	0,03	0,00	0,40
15	10,0	0,69	MULTIPLEX	15,95	**14,4**	**15,27**	0,09	0,00	0,00	**16,22**	0,05	0,02	0,00	0,00	0,02
53	10,0	0,19	MULTIPLEX	19,01	**5,8**	**1,69**	0,16	0,00	0,00	**2,04**	0,02	0,04	0,00	0,00	0,20
46	10,0	0,33	MULTIPLEX	20,39	**7,6**					0,14					0,00
47	10,0	0,34	MULTIPLEX	22,18	**6,7**					0,16					0,01
48	10,0	0,37	MULTIPLEX	20,37	**12,1**					0,08					0,03
49	10,0	0,35	MULTIPLEX	19,93	**8,9**					0,22					0,01
74	10,0	0,38	MULTIPLEX	20,36	**9,6**	0,07	0,23	0,00	0,00	0,55	0,00	0,00	0,00	0,00	0,04
75	10,0	0,10	MULTIPLEX	21,13	**8,7**	0,16	0,21	0,00	0,04	0,44	0,03	0,00	0,00	0,00	0,19
1	9,0	0,44	MULTIPLEX	17,98	**6,9**	0,27	0,00	0,00	0,00	0,53	0,04	0,00	0,00	0,00	0,21
6	10,0	0,50	MULTIPLEX	16,98	**7,9**					0,37					0,09
7	10,0	1,86	MULTIPLEX	14,44	**10,5**					0,15					0,01
8	10,0	0,93	MULTIPLEX	15,63	**8,9**					0,18					0,08
12	10,0	2,24	MULTIPLEX	14,06	**15,1**					0,05					0,02
14	10,0	0,35	MULTIPLEX	17,59	**10,2**					0,41					0,09
17	8,0	0,38	MULTIPLEX	18,65	**5,7**	0,22	0,62	0,14	0,06	0,61	0,00	0,20	0,10	0,00	0,01
18	8,0	0,64	MULTIPLEX	16,84	**7,2**	0,33	0,25	0,08	0,03	0,58	0,00	0,00	0,04	0,04	0,13
19	10,0	0,78	MULTIPLEX	16,44	**10,6**					0,29					0,01
20	10,0	0,70	MULTIPLEX	25,79	**9,3**					0,08					0,04
21	10,0	0,70	MULTIPLEX	17,03	**11,9**					0,08					0,00
22	8,0	0,37	MULTIPLEX	20,39	**11,1**					0,10					0,00
23	10,0	0,46	MULTIPLEX	17,94	**8,6**					0,14					0,04
26	10,0	0,35	MULTIPLEX	17,91	**10,2**					0,15					0,00
27	5,0	0,82	MULTIPLEX	16,99	**10,2**					0,31					0,05
29	10,0	1,66	MULTIPLEX	13,91	**14,7**					0,31					0,05
51	10,0	0,46	MULTIPLEX	19,52	**6,7**	0,47				0,23	0,02				0,03

Samples positive for one or more of the m*ESR1* mutations are in bold.

Table S6: LOD estimates of D538G after multiplex pre-amplification of 2 µL cfDNA

Sample	D538G input, VAF, %	D538G	
		pg in dPCR	VAF, %
Multi-pool	0.01	5286	0.09
Multi-pool	0.01	3798	0.65
Multi-pool	0.03	5296	0.36
Multi-pool	0.10	7211	0.10
Multi-pool	0.21	4265	0.26
Multi-pool	0.21	4169	0.30
Multi-pool	0.21	4473	0.69
Multi-pool	0.21	7350	0.27
Multi-pool	0.41	4223	0.78
Multi-pool	0.41	4153	0.56
Multi-pool	0.83	2065	1.71
Multi-pool	0.83	2281	1.24
Multi-pool	1.65	4128	1.33
Multi-pool	3.30	3662	4.38
Slope S			1.18
Standard error Sy			$3.5x10^{-3}$
LOD			0.89%

Supplemental Table 7: LOD estimates in HBDs after multiplex pre-amplification of 2 µL cfDNA

Sample	D538G		Y537C		Y537N		Y537S	
	pg in dPCR	VAF, %	pg in dPCR	VAF, %	pg in dPCR	VAF, %	pg in dPCR	VAF, %
HBD-1	10608	0.16	11265	0.21	10930	0.00	10857	0.04
HBD-2	8896	0.24	8742	0.10	8671	0.50	8481	0.61
HBD-3	8380	0.14	7195	0.24	7425	0.39	7980	0.07
HBD-4	7299	0.17	7236	0.06	6295	0.07	6776	0.00
HBD-5	12366	0.09	6978	0.16	6220	0.26	13410	0.00
HBD-6	6662	0.16	7445	0.21	7593	0.14	6579	0.00
HBD-7	7488	0.21	7361	0.33	6958	0.26	6613	0.06
HBD-8	7968	0.17	10571	0.21	9926	0.05	9065	0.03
HBD-9	33940	0.07	31390	0.09	32714	0.11	31385	0.01
HBD-10	15661	0.15	15552	0.09	10378	0.06	15676	0.00
HBD-11	12995	0.10	11014	0.12	10624	0.06	12312	0.02
HBD-12	13160	0.26	13839	0.20	12437	0.24	13941	0.02
HBD-13	9738	0.05	8297	0.13	8935	0.00	12364	0.06
HBD-14	15203	0.22	15786	0.10	13764	0.04	16339	0.03
HBD-15	11077	0.27	10024	0.07	10462	0.02	11624	0.00
HBD-16	9336	0.65	11387	0.12	9563	0.00	9769	0.05
HBD-17	11423	0.21	11839	0.07	12084	0.08	11798	0.00
HBD-18	11797	0.47	11526	0.18	8827	0.05	11630	0.00
HBD18_average		0.21		0.15		0.13		0.05
HBD18_SD		0.14		0.07		0.14		0.14
HBD18_2.58xSD		0.58		0.34		0.50		0.41
HBD18_MAX		0.65		0.33		0.50		0.61
HBD18_MAX + 2.58xSD		**1.02**		**0.52**		**0.87**		**0.97**

HBD, healthy blood donor

Table S8: Longitudinal monitoring of patients

Patient nr	Date of plasma sampling	Cohort	Days from start 1st line therapy	Therapy after start (BL)	ON sampling	Therapy before PD	endocrine therapy for MBC	Prior adjuvant therapy
12	25-10-2013	BL	15	AI (exemestane)				Tamoxifen + AI
12	29-11-2013	PD	50			AI (exemestane)		Tamoxifen + AI
16	18-07-2012	BL	1	AI (letrozole)				
16	11/12/2014	PD	847			AI (letrozole)		
23	24-09-2012	BL	11	Fulvestrant+dovitinib				AI + Anthracyclines + Taxanes
23	11/05/2012	PD	53			Fulvestrant + dovitinib		AI + Anthracyclines + Taxanes
25	10/12/2010	BL	8	AI (exemestane)				Tamoxifen + Anthracyclines+ Taxanes
25	29-03-2011	ON	176		AI (exemestane)			Tamoxifen + Anthracyclines+ Taxanes
25	26-05-2011	ON	234		AI (exemestane)			Tamoxifen + Anthracyclines+ Taxanes
25	24-01-2013	PD	843			AI (exemestane)		Tamoxifen + Anthracyclines+ Taxanes
25	13-05-2013	PD	952			Fulvestrant	AI (exemestane)	Tamoxifen + Anthracyclines+ Taxanes
29	16-09-2010	BL	34	AI (exemestane)				Tamoxifen + Anthracyclines
29	28-03-2011	PD	227			AI (exemestane)		Tamoxifen + Anthracyclines
42	09/08/2011	BL	10	AI (letrozole)				
42	23-01-2012	PD	147			AI (letrozole)		
51	21-08-2012	BL	28	AI (anastrozole)				
51	28-02-2014	PD	583			Tamoxifen	AI (anastrozole)	
57	11/10/2009	BL	63	AI (exemestane)				Tamoxifen
57	01/07/2010	ON	121		AI (exemestane)			Tamoxifen
60	12/08/2011	BL	21	Tamoxifen				
60	01/12/2012	ON	56		Tamoxifen			
60	09/03/2012	ON	291		Tamoxifen			
77	11/06/2012	BL	40	AI (anastrozole)				Tamoxifen
77	20-02-2014	PD	482			AI (exemestane) + everolimus	AI (anastrozole)	Tamoxifen
79	22-04-2010	BL	10	Tamoxifen				
79	28-05-2010	ON	46		Tamoxifen			
93	16-80-2012	BL	3	AI (anastrozole)				Tamoxifen
93	26-10-2012	PD	74			AI (anastrozole)		Tamoxifen
103	30-06-2011	BL	9	AI (anastrozole)				
103	08/08/2011	ON	48		AI (anastrozole)			•
104	16-04-2013	PD	932			AI (anastrozole)		•
104	21-10-2013	ON	1120		Tamoxifen		AI (anastrozole)	•
122	29-06-2010	BL	18	AI (letrozole)				• Anthracyclines + Tamoxifen
122	08/03/2010	ON	53		AI (letrozole)			• Anthracyclines + Tamoxifen
131	06/01/2012	ON	65		AI (letrozole)			• Anthracyclines + Tamoxifen
131	30-04-2013	PD				AI (letrozole)		• Anthracyclines + Tamoxifen
174	20-08-2010	BL	23	AI (anastrozole)				•
174	03/01/2012	PD	582			Anthracyclines + Tamoxifen	AI (anastrozole)	•

Continued.					VAF in cfDNA				
Patient nr	**Date of plasma sampling**	**Cohort**	**cfDNA ng/µL**	***ESR1* mutated at 1%**	**D538G**	**Y537C**	**Y537N**	**Y537S**	**input in dPCR (pg)**
12	25-10-2013	BL	0,2	no	0,67	0,30	0,20	0,06	7244
12	29-11-2013	PD	2,24	no	0,47	0,14	0,06	0,00	11140
16	18-07-2012	BL	1,55	no	0,33	0,20	0,02	0,00	15231
16	11/12/2014	PD	0,84	**yes**	0,18	0,29	**6,39**	0,00	14139
23	24-09-2012	BL	0,56	no	0,00	0,04	0,35	0,33	6115
23	11/05/2012	PD	0,46	no	0,02	0,19	0,00	0,00	8767
25	10/12/2010	BL	0,3	no	0,30	0,18	0,17	0,00	9591
25	29-03-2011	ON	0,44	no	0,05	0,05	0,05	0,05	10282
25	26-05-2011	ON	0,46	no	0,10	0,82	0,30	0,30	12247
25	24-01-2013	PD	0,18	**yes**	**25,16**	0,13	**1,61**	0,01	9519
25	13-05-2013	PD	0,56	**yes**	**8,73**	0,10	**6,26**	0,02	11502
29	16-09-2010	BL	0,51	no	0,25	0,42	0,05	0,15	7876
29	28-03-2011	PD	0,37	no	0,16	0,04	0,04	0,14	11562
42	09/08/2011	BL	0,1	no	0,14	0,27	0,13	0,00	6767
42	23-01-2012	PD	2,1	no	0,19	0,27	0,06	0,05	11351
51	21-08-2012	BL	0,32	no	0,10	0,04	0,00	0,00	10514
51	28-02-2014	PD	0,49	**yes**	0,27	0,11	0,00	**3,25**	8119
57	11/10/2009	BL	0,29	no	0,17	0,19	0,19	0,19	7394
57	01/07/2010	ON	0,39	no	0,35	0,48	0,00	0,00	7148
60	12/08/2011	BL	0,32	**yes**	**1,90**	0,40	0,17	0,03	18564
60	01/12/2012	ON	1,35	no	0,01	0,01	0,01	0,01	14413
60	09/03/2012	ON	0,49	no	0,29	0,11	0,03	0,00	13326
77	11/06/2012	BL	0,33	no	0,00	0,07	0,00	0,12	9594
77	20-02-2014	PD	1,89	no	0,73	0,42	0,00	0,04	12172
79	22-04-2010	BL	0,7	no	0,16	0,41	0,10	0,10	18464
79	28-05-2010	ON	1,13	no	0,38	0,08	0,08	0,08	30050
93	16-80-2012	BL	0,89	no	0,65	0,16	0,05	0,00	7758
93	26-10-2012	PD	0,39	no	0,21	0,05	0,03	0,03	11172
103	30-06-2011	BL	0,47	no	0,17	0,04	0,16	0,00	10731
103	08/08/2011	ON	0,45	no	0,35	0,07	0,07	0,07	18606
104	16-04-2013	PD	0,54	no	0,05	0,05	0,05	0,05	13014
104	21-10-2013	ON	0,34	no	0,24	0,06	0,06	0,06	9315
122	29-06-2010	BL	0,4	no	0,15	0,02	0,18	0,00	11000
122	08/03/2010	ON	0,33	no	0,04	0,04	0,04	0,04	15866
131	06/01/2012	ON	0,43	no	0,07	0,07	0,07	0,07	13738
131	30-04-2013	PD	0,82	no	0,17	0,20	0,24	0,00	11282
174	20-08-2010	BL	0,55	no	0,13	0,20	0,04	0,00	11113
174	03/01/2012	PD	0,35	no	0,14	0,08	0,05	0,00	8336

Samples positive for one or more of the m*ESR1* mutations are in bold.

BL, baseline; OT, on-treatment; PD, progressive disease

Chapter 7

The prognostic and predictive value of *ESR1* fusion gene transcripts in primary breast cancer[2]

Vitale SR, Ruigrok-Ritstier K, Timmermans AM, Foekens R, Trapman-Jansen AMA., Beaufort CM, Vigneri P, Sleijfer S, Martens JWM, Sieuwerts AM, Jansen MPHM

BMC Cancer 22, 165 (2022).
doi: 10.1186/s12885-022-09265-1

[2] The chaper is a slightly modified version of the paper which is published.

Abstract

In breast cancer (BC), recurrent fusion genes of estrogen receptor alpha (*ESR1*) and *AKAP12*, *ARMT1* and *CCDC170* have been reported. In these gene fusions the ligand binding domain of ESR1 has been replaced by the transactivation domain of the fusion partner constitutively activating the receptor. As a result, these gene fusions can drive tumor growth hormone independently as been shown in preclinical models, but the clinical value of these fusions have not been reported. Here, we studied the prognostic and predictive value of different frequently reported *ESR1* fusion transcripts in primary BC.

We evaluated 732 patients with primary BC (131 *ESR1*-negative and 601 *ESR1*-positive cases), including two ER-positive BC patient cohorts: one cohort of 322 patients with advanced disease who received first-line endocrine therapy (ET) (predictive cohort), and a second cohort of 279 patients with lymph node negative disease (LNN) who received no adjuvant systemic treatment (prognostic cohort). Fusion gene transcript levels were measured by reverse transcriptase quantitative PCR. The presence of the different fusion transcripts was associated, in uni- and multivariable Cox regression analysis taking along current clinic-pathological characteristics, to progression free survival (PFS) during first-line endocrine therapy in the predictive cohort, and disease- free survival (DFS) and overall survival (OS) in the prognostic cohort.

The *ESR1-CCDC170* fusion transcript was present in 27.6% of the *ESR1*-positive BC subjects and in 2.3% of the ESR1-negative cases. In the predictive cohort, none of the fusion transcripts were associated with response to first-line ET. In the prognostic cohort, the median DFS and OS were respectively 37 and 93 months for patients with an *ESR1-CCDC170* exon 8 gene fusion transcript and respectively 91 and 212 months for patients without this fusion transcript. In a multivariable analysis, this *ESR1-CCDC170* fusion transcript was an independent prognostic factor for DFS (HR) (95% confidence interval (CI): 1.8 (1.2–2.8), P=0.005) and OS (HR (95% CI: 1.7 (1.1–2.7), P=0.023).

Our study shows that in primary BC only *ESR1-CCDC170* exon 8 gene fusion transcript carries prognostic value. None of the *ESR1* fusion transcripts, which are considered to have constitutive ER activity, was predictive for outcome in BC with advanced disease treated with endocrine treatment.

Introduction

The estrogen receptor (ER) plays a key role in cellular growth and tumor development in a large fraction of breast cancers. As a result, endocrine therapy has been and still is a successful treatment in patients with *ESR1*-positive (*ESR1*+) breast cancers (BC) [1]. However, in the metastatic setting, nearly half of the patients are de novo resistant to endocrine therapy while the remaining cases acquire resistance over time [2, 3]. One of the primary characterized mechanisms of acquired resistance to endocrine therapy is the acquisition of mutations within the ligand-binding domain (LBD) of the estrogen receptor alpha gene (*ESR1*) activating the receptor constitutively thereby rendering tumor cells less dependent on estrogen [4-7]. Another mechanism that lead to less estrogen dependency of BC cells is the occurrence of ESR1 fusion proteins. Through analysis of RNA-sequencing data in breast cancer, recurrent intragenic fusions of 5' end of ESR1 and the 3' ends of *AKAP12*, *ARMT1* or *CCDC170* amongst other genes have been identified [8-13]. *AKAP12*, *ARMT1*, and *CCDC170* genes together with *ESR1* gene were selected for our evaluation, because they all were located at the 6q25.1 locus within 1Mb distance [14] and fusions between the two non-coding 5' exons of *ESR1* with the 3' ends of *CCDC170*, *AKAP12* and *ARMT1*, upstream of *ESR1*, were identified in patients resistant to endocrine treatment [9, 10].

Gene fusions were preferentially detected in high-grade disease and/or endocrine-resistant forms of ESR1+ BC [10, 13]. Particularly, an enrichment of *ESR1-CCDC170* fusion was previously reported in HER-positive patients (luminal A 9%, luminal B 3-8% and HER2 3.1%) and was correlated with a worse clinical outcome after endocrine therapy[9, 15, 16]. The *ESR1-AKAP12* fusion was identified in 6.5% breast cancer that were resistant to letrozole aromatase inhibitor treatment [17].The novel fusion *ESR1-ARMT1* was instead detected in a HER2-negative patient with luminal A-like subtype [16] and in a breast cancer patient who had not received endocrine therapy [18]. Moreover, a recently study based on molecular characterization of luminal breast cancer in African American women reported the fusions at a frequency of 11% for *ESR1-CCDC170*, 8% for *ESR1-AKAP12* and 6% for *ESR1-ARMT1*[19]. Despite the diversity among these fusions, they share a common structure retaining the hormone-independent transactivation domain as well as the DNA-binding domain whereas their ligand-binding domain is lost and replaced with a functional (transactivating) domain of the fusion partner, suggesting a pathological impact in *ESR1*+ BC [13]. However, the clinical significance of these fusions has not yet been properly addressed in uniform and well annotated cohorts.

In this study, we explored the occurrence of fusion transcripts of three of the most commonly reported fusion partners of *ESR1* (i.e. *CCDC170*, *AKAP12* and *ARMT1*) and determined the associations of their presence with clinical outcome in a cohort of 732 breast cancer patients allowing us to investigate their predictive value for endocrine treatment failure as well as their prognostic value.

Materials and methods

Study Cohorts

The protocol to study biological markers associated with disease outcome was approved by the medical ethics committee of the Erasmus Medical Centre Rotterdam, The Netherlands (MEC 02.953) and was performed in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in The Netherlands (https://www.federa.org/codes-conduct). The use of coded left-over material for scientific purposes and, therefore, for the greater good, does not require informed consent according to Dutch law and the new European general data protection regulation (GDPR).

In this retrospective study (see **Figure 1A** for the consort diagram of the study), female patients were included, who underwent surgery for invasive primary breast cancer between 1980 and 2000 in the Netherlands. A further selection criterion was no previously diagnosed cancers with the exception of basal cell carcinoma or stage Ia/Ib cervical cancer. Within this study, only data from sections of primary tumors with at least 30% invasive tumor cells were included. The details of tissue processing, RNA isolation, cDNA synthesis and QC of this cohort have been described previously [20, 21]. Tumor grade was assessed according to standard procedures at the time of inclusion. For the classification of patients' RNA samples regarding expression of the estrogen and progesterone receptors, as well as the human epidermal growth factor receptor 2 (HER2) amplification status, reverse transcriptase quantitative PCR (RT-qPCR) was used with cut-offs previously described by us [20, 21].

The total cohort consisted of 732 patients with primary breast cancer (131 *ESR1*-negative and 601 *ESR1*-positive cases) (**Figure 1B**). The clinical relevance of the gene fusion transcripts was evaluated in a predictive and a prognostic cohort of *ESR1*+ BC patients.

The predictive cohort consisted of 322 breast cancer patients with *ESR1*+ primary tumors of which 235 patients received tamoxifen (40 mg daily) and 87 patients an aromatase inhibitor (AI:

anastrozole, letrozole, exemestane [22]) as a 1st-line treatment for recurrent disease. Clinical response to tamoxifen therapy was defined as previously described [20, 23]. The prognostic cohort included primary tumors from 279 lymph node negative (LNN) *ESR1*+ BC patients who had not received any systemic (neo) adjuvant therapy. Of note, 122 of these LNN ESR1+ patients were also included in the predictive cohort. Clinicopathological characteristics of each of these 2 cohorts are described in **Table 1**. Association of *ESR1* fusions with clinical parameters of patients enrolled in the predictive cohort and in the prognostic cohort are reported in **Table 2** and **Table 3**, respectively**.**

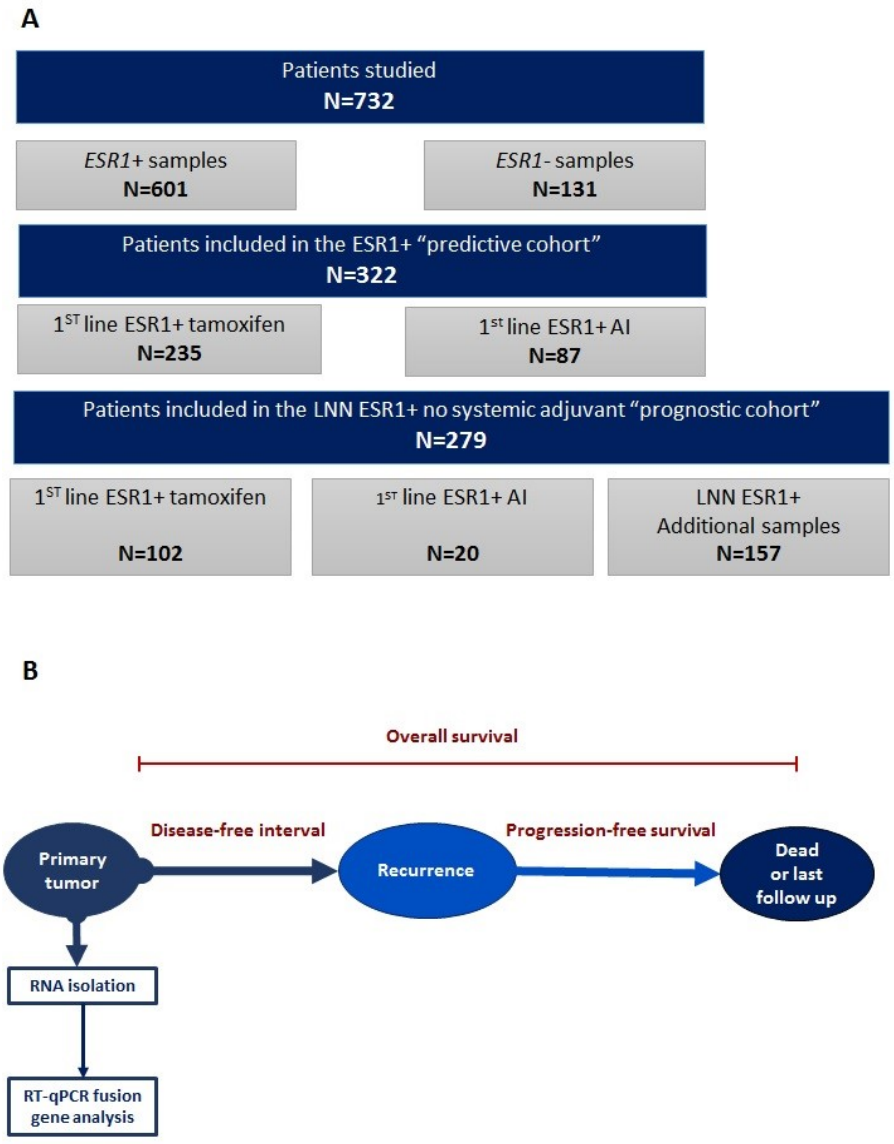


Figure 1. Overview of the study and selection of available patients.

A) Flow diagram of the study; **B**) Workflow of processing samples: fusion gene mRNA levels were measured in 322 ER-positive primary tumors (predictive cohort) by quantitative reverse transcriptase PCR (RT-qPCR). All patients in this cohort were hormone-naïve and all experienced a disease recurrence and subsequently received 1st line endocrine therapy. The association of the presence of *ESR1* fusion genes in the primary tumor progression-free survival (PFS) after start with 1st line tamoxifen (n=235) or aromatase inhibitors (n=87), were evaluated. Similarly, disease free interval (DFS) and overall survival (OS) were investigated in 279 lymph node negative ER-positive breast cancer patients (prognostic cohort) who had not received any (neo)adjuvant systemic therapy. ESR1: Estrogen Receptor 1 gene; AI: Aromatase Inhibitor; LNN: Lymph node negative; ER: Estrogen Receptor; RT-qPCR: Quantitative reverse transcriptase PCR.

Table 1. Clinicopathological characteristics of *ESR1*-positive breast cancer patient cohorts

	Predictive Cohort (N= 322)		Prognostic Cohort
	Tamoxifen	Aromatase inhibitors	Lymph node negative (LNN)
Total	235	87	279
Median age (range)	61 (29-90)	66 (35-86)	55 (26-85)
Menopausal Status:			
Premenopausal	60	4	120
Postmenopausal	175	83	159
Surgery:			
Lumpectomy	87	8	178
Ablation	148	79	101
Adjuvant hormonal therapy:			
no	235	17	279
yes	0	70	0
Adjuvant chemotherapy:			
no	198	69	279
yes	37	18	0
Lymph node status:			
negative	102	20	279
positive	81	50	0
not applicable (M1)	52	17	0
Distant metastasis:			
yes	235	87	165
no	0	0	114
Disease -Free Interval:			
<1 year	59	13	20
1-3 year	108	29	71
>3 year	68	45	188
Median Follow-up time (in months):			
after surgery	62 (3-272)	103 (7-295)	93 (5-337)
after start therapy	30 (1-208)	45 (2-108)	
PR status*:			
Positive	187	72	217
Negative	48	15	62
HER2 status*:			
Amplified	32	10	43
Not amplified	203	77	236

*as measured by RT-qPCR<; *ESR1*: estrogen receptor alpha; LNN: lymph node negative disease; M1: methastatic stage 1; ND: data not available; PR: progesterone receptor; HER2: human epidermal growth factor receptor 2; CCDC170: coiled-coil domain containing 170; RT-qPCR: Quantitative Real-Time Polymerase Chain Reaction.

Table 2. Association of *ESR1* fusions with clinical parameters in the predictive cohort

		Predictive Endocrine Therapy Cohorts											
Parameters*	**n**	***at least one ESR1-CCDC170 (exon 2 to 8) fusion***		***P-Value***	***ESR1-CCDC170 (exon 2) fusion***		***PValue***	***ESR1-CCDC170 (exon 8) fusion***		***P-Value***	***ESR1-AKAP12***		***P-Value***
		n	**%**		**n**	**%**		**n**	**%**		***n***	***%***	
All patients	322	89	27.6%		50	15.5%		51	15.8%		13	4.0%	
Age at start 1st line treatment (years)													
≤ 50	63	19	30.2%	*0.63*	12	19.0%	*0.62*	8	12.7%	***0.029***	1	1.6%	*0.36*
>50-≤ 70	161	37	23.0%		23	14.3%		24	14.9%		7	4.3%	
>70	98	33	33.7%		15	15.3%		19	19.4%		5	5.1%	
Menopausal status at start of 1st line treatment													
Premenopausal	64	17	26.6%	*0.82*	10	15.6%	*0.99*	8	12.5%	*0.41*	1	1.6%	*0.26*
Postmenopausal	258	72	28.0%		40	15.6%		43	16.7%		12	4.7%	
Surgery type													
Lumpectomy	95	25	26.3%	*0.79*	14	14.7%	*0.90*	15	15.8%	*0.83*	2	2.1%	*0.89*
Ablation	227	42	24.9%		24	14.2%		25	14.8%		4	2.4%	
Radiotherapy													
No	105	30	28.6%	*0.33*	20	19.0%	*0.08*	16	15.2%	*0.98*	2	1.9%	*0.74*
Yes	159	37	23.3%		18	11.3%		24	15.1%		4	2.5%	
Nodal status													
No lymph nodes	122	33	27.0%	*0.88*	19	15.6%	*0.99*	20	16.4%	*0.95*	4	3.3%	*0.2*
Positive lymph nodes	131	38	29.2%		21	16.2%		22	16.9%		9	6.9%	
Tumor outside lymph nodes	53	15	28.3%		8	15.1%		7	13.2%		0	0.0%	
Not applicable (M1)	16	3	18.8%		2	12.5%		2	12.5%		0	0.0%	
Pathological Tumor classification													
pT1	85	22	25.9%	*0.60*	13	15.3%	*0.21*	14	16.5%	*0.90*	2	2.4%	*0.36*
pT2 + unknown	186	50	26.9%		25	13.4%		30	16.1%		10	5.4%	
pT3+pT4	51	17	33.3%		12	23.5%		7	13.7%		1	2.0%	
Tumor grade													
Poor	167	45	28.1%	*0.36*	27	16.9%	*0.60*	27	16.9%	*0.60*	7	4.4%	*0.078*
Unknown	81	18	22.2%		10	12.3%		10	12.3%		0	0.0%	

Moderate/Good	74	24	32.4%		13	17.6%		13	17.6%		5	6.8%	
Tumor cell content													
30-49%	27	7	25.9%	*0.96*	4	14.8%	*0.99*	2	7.4%	*0.25*	2	7.4%	*0.63*
50-70%	98	28	28.6%		15	15.3%		13	13.3%		4	4.1%	
>70%	197	54	27.4%		31	15.7%		36	18.3%		7	3.6%	
Hormone/ growth factor status (RT-qPCR)*													
ESR1-negative	0	0			0			0			0		
ESR1-positive	322	89	27.6%		50	15.5%		51	15.8%		13	4.0%	
PR-negative	63	18	28.6%	*0.87*	11	17.5%	*0.65*	11	17.5%	*0.70*	6	9.5%	***0.014***
PR-positive	259	71	27.5%		39	15.1%		40	15.5%		7	2.7%	
HER2 non-amplified	280	77	27.6%	*0.63*	44	15.8%	*0.85*	45	16.1%	*0.81*	13	4.7%	*0.16*
HER2 amplified	42	12	29.3%		6	14.6%		6	14.6%		0	0.0%	
CCDC170 negative	31	5	16.1%	*0.13*	2	6.5%	*0.15*	4	12.9%	*0.62*	0	0.0%	*0.23*
CCDC170 positive	287	83	28.9%		47	16.4%		47	16.4%		13	4.5%	
Adjuvant endocrine therapy													
No	252	66	26.2%	*0.24*	38	15.1%	*0.64*	36	14.3%	*0.13*	7	2.8%	***0.030***
Yes (AI cohort only)	69	23	33.3%		12	17.4%		15	21.7%		6	8.7%	
Adjuvant chemotherapy													
No	267	76	28.5%	*0.47*	40	15.0%	*0.55*	45	16.9%	*0.27*	12	4.5%	*0.36*
Yes	55	13	23.6%		10	18.2%		6	10.9%		1	1.8%	
Disease-free interval													
≤ 1 year disease-free	72	23	31.9%	*0.47*	14	19.4%	*0.62*	12	16.7%	*0.99*	2	2.8%	*0.45*
1-3 years disease-free	137	37	27.0%		20	14.6%		20	14.6%		8	5.8%	
>3 years disease-free	113	29	25.7%		16	14.2%		19	16.8%		3	2.7%	
Dominant site of metastasis													
Local regional	29	10	34.5%	*0.51*	7	24.1%	*0.32*	4	13.8%	*0.36*	0	0.0%	*0.40*
Bone	159	40	25.2%		25	15.7%		21	13.2%		6	3.8%	
Other distant metastasis	130	38	29.2%		17	13.1%		25	19.2%		7	5.4%	
Response type													

	Complete response	11	3	27.3%	*0.87*	2	18.2%	*0.73*	1	9.1%	*0.29*	0	0.0%	*0.46*
	Partial response	39	9	23.1%		3	7.7%		6	15.4%		2	5.1%	
	Stable disease over 6 months (SD > 6m)	115	32	27.8%		16	13.9%		23	20.0%		1	0.9%	
	Stable disease for 6 months or less (SD ≤ 6m)	13	2	15.4%		2	15.4%		1	7.7%		0	0.0%	
	Progressive disease (PD)	83	20	24.1%		14	16.9%		8	9.6%		3	3.6%	
Response type														
	No response	96	22	22.9%	*0.50*	16	16.7%	*0.38*	9	9.4%	*0.05*	3	3.1%	*0.50*
	Response	165	44	26.7%		21	12.7%		30	18.2%		3	1.8%	

ESR1: estrogen receptor alpha; CCDC170: coiled-coil domain containing 170; AKAP12: A-Kinase Anchoring Protein 12 gene; *ESR1-CCDC170*: *ESR1-CCDC170* gene fusion; *ESR1-AKAP12*: *ESR1-AKAP12* gene fusion; M1: methastatic stage 1; pT: primary tumor; pT1: small primary tumor (tumour is 2 centimetre across or less); pT2: tumour more than 2 centimetres but no more than 5 centimetres across; pT3: T3 tumour bigger than 5 centimetres across; pT4: tumor with phatological stage; RT-qPCR: Quantitative Real-Time Polymerase Chain Reaction; PR: PR: progesterone receptor; HER2: human epidermal growth factor receptor; AI: aromatase inhibitors;SD: standar deviation; PD: progressive disease. Statistically significant differences are indicated in bold. *due to unknown data, numbers do not add up to 322.

Table 3. Associations of *ESR1* fusions with clinical parameters in prognostic clinical cohort

		LNN ESR+ Prognostic cohort											
Parameters	**n**	***at least one ESR1-CCDC170 (exon 2 to 8) fusion***		***P-value***	***ESR1-CCDC170 (exon 2) fusion***		***P-value***	***ESR1-CCDC170 (exon 8) fusion***		***P-value***	***ESR1-AKAP12***		***P-value***
		n	**%**		**n**	**%**		**n**	**%**		***n***	***%***	
All patients	279	66	23.7%		33	11.8%		39	14.0%		5	1.8%	
Age at primary surgery													
≤ 40 years	29	6	20.7%	***0.001***	4	13.8%	*0.38*	4	13.8%	*0.26*	1	3.4%	*0.27*
41-50 years	81	11	13.6%		5	6.2%		5	6.2%		0	0.0%	
51-70 years	125	36	28.8%		16	12.8%		21	16.8%		3	2.4%	
>70 years	44	17	38.6%		8	18.2%		9	20.5%		1	2.3%	
Menopausal status													
Premenopausal	120	19	15.8%	***0.002***	10	8.3%	*0.12*	11	9.2%	***0.044***	1	0.8%	*0.29*
Postmenopausal	159	51	32.1%		23	14.5%		28	17.6%		4	2.5%	
Surgery type													
Lumpectomy	178	44	24.7%	*0.85*	19	10.7%	*0.43*	25	14.0%	*0.97*	4	2.2%	*0.45*
Ablation	101	26	25.7%		14	13.9%		14	13.9%		1	1.0%	
Radiotherapy													
No	84	24	28.6%	*0.38*	14	16.7%	*0.10*	12	14.3%	*0.92*	1	1.2%	*0.62*
Yes	195	46	23.6%		19	9.7%		27	13.8%		4	2.1%	
Nodal status													
No lymph nodes	279	70	25.1%		33	11.8%		39	14.0%		5	1.8%	
Positive lymph nodes	0	0			0			0			0		
Tumor outside lymph nodes	0	0			0			0			0		
Pathological Tumor classification													
pT1	151	34	22.5%	*0.28*	17	11.3%	*0.61*	16	10.6%	*0.08*	2	1.3%	*0.1*
pT2 + unknown	119	32	26.9%		14	11.8%		20	16.8%		2	1.7%	
pT3+pT4	9	4	44.4%		2	22.2%		3	33.3%		1	11.1%	

Tumor grade												
Poor	131	36	27.5%	*0.60*	21	16.0%	*0.06*	21	16.0%	*0.56*	3	2.3%	*0.84*
Unknown	81	20	24.7%		9	11.1%		11	13.6%		1	1.2%	
Moderate/Good	67	14	20.9%		3	4.5%		7	10.4%		1	1.5%	
Tumor cell content												
30-49%	31	9	29.0%	*0.82*	6	19.4%	*0.38*	4	12.9%	*0.86*	1	3.2%	*0.81*
50-70%	69	16	23.2%		7	10.1%		11	15.9%		1	1.4%	
>70%	179	45	25.1%		20	11.2%		24	13.4%		3	1.7%	
Hormone/ growth factor status (RT-qPCR)*												
ESR1 negative	0	0			0			0			0		
ESR1 positive	279	70	25.1%		33	11.8%		39	14.0%		5	1.8%	
PR negative	62	16	25.8%	*0.88*	9	14.5%	*0.46*	8	12.9%	*0.78*	2	3.2%	*0.93*
PR positive	217	54	24.9%		24	11.1%		31	14.3%		3	1.4%	
HER2 non-amplified	236	62	26.6%	*0.15*	29	12.4%	*0.30*	34	14.6%	*0.61*	4	1.7%	*0.78*
HER2 amplified	43	7	16.3%		3	7.0%		5	11.6%		1	2.3%	
CCDC170 negative	27	4	15.4%	*0.23*	2	7.7%	*0.49*	3	11.5%	*0.70*	0	0.0%	*0.47*
CCDC170 positive	252	66	26.2%		31	12.3%		36	14.3%		5	2.0%	
Disease-free interval												
≤ 1 year disease-free	20	7	35.0%	***0.011***	2	10.0%	*0.08*	4	20.0%	***0.006***	0	0.0%	*0.57*
1-3 years disease-free	71	18	25.4%		10	14.1%		14	19.7%		2	2.8%	
>3 years disease-free	188	45	23.9%		21	11.2%		21	11.2%		3	1.6%	
Adjuvant endocrine therapy												
No	279	66	23.7%		33	11.8%		39	14.0%		5	1.8%	
Yes	0	0			0			0			0		
Adjuvant chemotherapy												
No	279	66	23.7%		33	11.8%		39	14.0%		5	1.8%	
Yes	0	0			0			0			0		

ESR1: estrogen receptor alpha; CCDC170: coiled-coil domain containing 170; AKAP12: A-Kinase Anchoring Protein 12 gene; *ESR1-CCDC170*: *ESR1-CCDC170* gene fusion; *ESR1-AKAP12*: *ESR1-AKAP12* gene fusion; pT: primary tumor; pT1: small primary tumor (tumour is 2 centimetre across or less); pT2: tumour more than 2 centimetres but no more than 5 centimetres across; pT3: T3 tumour bigger than 5 centimetres across; pT4: tumor with phatological stage; RT-qPCR: Quantitative Real-Time Polymerase Chain Reaction; PR: PR: progesterone receptor; HER2: human epidermal growth factor receptor. Statistically significant differences are indicated in bold.

RNA isolation and RT-qPCR

Total RNA isolation from human breast cancer tissue, breast cancer cell line models and quality control were performed as previously described [14]. Next, cDNA was generated by a cycle at 48°C for 30 minutes with RevertAid H-minus (Applied Biosystems, Carlsbad, CA), according to the manufacturer's instructions. The cDNA was then pre-amplified for specific genes as previously described [14]. Briefly, 2 µL of cDNA (0.1 to 1 ng/ µL) was subject to a pre-amplification of 15 cycles using a multiple loci target-specific amplification for *ESR1* fusions with *AKAP12*, *ARMT1* and *CCDC170* and two reference genes, the Epithelian Cell Adhesion Molecule (*EPCAM)* and the Hypoxanthine Phosphoribosyltransferase 1 (*HPRT1*), with TaqMan PreAmp Master Mix (Applied Biosystems), as recommended by the manufacturer. Pre-amplified products were then diluted 12-fold in LoTE buffer (3 mM Tris-HCl/0.2 mM EDTA, pH 8.0) prior to downstream analysis. Next, 5 µL diluted pre-amplified samples were subjected to a TaqMan probe based real-time quantitative PCR (qPCR) for each gene combination, according to the manufacturer's instructions, in a MX3000P Real-Time PCR System (Agilent, Santa Clara, CA). The average expression of *HPRT1* and the epithelial marker *EPCAM* was used as reference to control RNA quality and calculate the expression levels of target genes, as previously described [14]. Only those samples with a ΔCq >25 relative to the two reference genes were used for further evaluation of gene fusions, as previously described [18-20]. **Table S1** describes the primer sets used in the pre-amplification combination, as well as the Taqman qPCR used to quantify the fusions and reference genes. For *ESR1-CCDC170* fusion transcripts, the variants in which exon 2 of *ESR1* is fused to the coding region (exon 2 to 11) of *CCDC170* were examined (E2-E2, E2-E3, E2-E4, E2-E5, E2–E6, E2–E7, E2–E8, E2–E10 and E2-E11). Samples with a ΔCq >25 relative to the reference genes were afterwards validated by MultiNA analysis (Shimadzu Europe, Duisburg, Germany). Only those samples with a MultiNA fusion product of the expected size were considered positive for the fusion transcripts **(Table S2).** The detection of *ESR1-CCDC170* fusion transcripts with RT-qPCR and MultiNA analysis was verified and confirmed in a set of fusion-positive reported breast cancer cell lines (**Figure S1, S2 and S3).**

Statistical analysis

All data were entered in SPSS version 24 (IBM Corp., Armonk, NY, USA) to generate the tables and perform the statistical analyses. For contingency tables, the Pearson Chi-Square Test was used. All *P*-values are 2-sided and $P<0.05$ was considered statistically significant.

Results

Association of *ESR1* with its *CCDC170*, *AKAP12* and *ARMT1* fusion partner

The presence of the *ESR1* fusions with *AKAP12*, *ARMT1* and *CCDC170 (exon 2 to exon 11)* was evaluated in breast cancer tissue samples from 732 breast cancer patients. Fusion transcripts were predominantly detected in the *ESR1*+ population, with *CCDC170*, *AKAP12* or *ARMT1* fusion transcripts observed in 27.6%, 4.04% and 1.4% of the ER-positive cases respectively, and seen in 2.3%, 0.76% and 0% of the *ESR1*- cases respectively (*P* <0.001, Fisher's exact test two tailed **Table 4** and **Table S3**). In ER-positive tumors, full length *ESR1* and *CCDC170* mRNA levels were strongly correlated (R^2=0.31, *P* <0.0001) (**Figure S4A**) and transcript levels of both were significantly higher in the group of samples with an *ESR1-CCDC170* fusion transcript when compared to the group without [Student T-Test *P* = 0.0316 and 0.0001, respectively (**Figure S4B**)].

Table 4. Prevalence of *ESR1* fusions in the different analyzed cohorts.

			At least one *ESR1-CCDC170* (exon 2 to 8) fusion				***ESR1-CCDC170* exon 2**			***ESR1-CCDC170* exon 8**			***ESR1_AKAP12***				***ESR1-ARMT1***			
	Total Count		**no**	**yes**	**%**	**% of total count**	**no**	**yes**	**%**	**no**	**yes**	**%**	**no**	**yes**	**%**	**% of total count**	**no**	**yes**	**%**	**% of total count**
All samples studied	**732***	ESR1 negative	128	3	2.3%	**29.9%**	128	3	2.29%	130	1	0.76%	130	1	0.76%	**4.8%**	131	0	0.00%	**1.4%**
		ESR1 positive	435	166	27.6%		513	88	14.64%	504	97	16.1%	574	27	4.04%		592	9	1.4%	
1st line Tamoxifen	**235**	ESR1 negative	0	0	0%	**24.7%**	0	0	0%	0	0	0%	0	0	0%	**2.6%**	0	0	0%	**0.4%**
		ESR1 positive	177	58	24.7%		204	31	13.19%	201	34	14.47%	229	6	2.55%		234	1	0.43%	
1st line AI	**87**	ESR1 negative	0	0	0%	**35.6%**	0	0	0%	0	0	0%	0	0	0%	**8.0%**	0	0	0%	**3.4%**
		ESR1 positive	56	31	35.6%		68	19	21.84%	70	17	19.54%	80	7	8.05%		84	3	3.45%	
Predictive cohort	**322**	ESR1 negative	0	0	0%	**27.6%**	0	0	0%	0	0	0%	0	0	0%	**4.0%**	0	0	0%	**1.2%**
		ESR1 positive	233	89	27.6%		272	50	15.53%	271	51	15.84%	309	13	4.04%		318	4	1.24%	
Primary LNN cohort investigated	**369**	ESR1 negative	87	3	3.3%	**26.9%**	87	3	3.33%	89	1	1.11%	89	1	1.11%	**2.9%**	90	0	0.0%	**1.08%**
		ESR1 positive**	213	66	23.7%		246	33	11.83%	240	39	13.98%	274	5	1.79%		276	3	1.08%	
Normal breast tissue of breast cancer patients	**36**	ESR1 negative	0	0	0%	**66.7%**	0	0	0%	0	0	0%	0	0	0%	**0.0%**	0	0	0%	**0.0%**
		ESR1 positive	12	24	66.7%		18	18	50.0%	23	13	36.1%	36	0	0.0%		36	0	0.0%	
	16	ESR1 negative	0	0	0%	**25.0%**	0	0	0%	0	0	0%	0	0	0%	**0.0%**	0	0	0%	**0.0%**

Tissue of breast fibroadenoma's		ESR1 positive	12	4	25.0%		16	0	0.0%	12	4	25.0%	16	0	0.0%		16	0	0.0%	
Tissue of breast DCIS	**13**	ESR1 negative	0	0	0%	**7.7%**	0	0	0%	0	0	0%	0	0	0%	**0.0%**	0	0	0%	**0.0%**
		ESR1 positive	12	1	7.7%		13	0	0.0%	13	0	0.0%	13	0	0.0%		13	0	0.0%	
Normal breast tissue of healthy women	**10**	ESR1 negative	0	0	0%	**10.0%**	0	0	0%	0	0	0%	0	0	0%	**0.0%**	0	0	0%	**0.0%**
		ESR1 positive	9	1	10.0%		9	1	10.0%	10	0	0.0%	10	0	0.0%		10	0	0.0%	

ESR1: estrogen receptor alpha; CCDC170: coiled-coil domain containing 170; *AKAP12*: A-Kinase Anchoring Protein 12 gene; ARMT1: Acidic Residue Methyltransferase 1*; ESR1-CCDC170*: *ESR1-CCDC170* gene fusion; *ESR1-AKAP12*: *ESR1-AKAP12* gene fusion; *ESR1-ARMT1*: *ESR1-ARMT1* gene fusion; 1st: first line treatment; LNP: lymph node positive disease; LNN: lymph node negative disease; DCIS: Ductal carcinoma in situ. Statistically significant differences are indicated in bold.* not all patients were further analyzed; **: LNN ESR1+ prognostic cohort included into the study.

Prevalence of *ESR1* fusion genes in normal mammary tissue, benign lesions and carcinoma in situ of the breast

While *AKAP12* and *ARMT1* fusion transcripts were not found in 36 non-malignant breast tissues taken at a distance of the primary tumor, *ESR1-CCDC170* fusion transcripts were detected in 66.7% of these normal breast tissues of patients with diagnosed breast cancer (**Table 4)**. Note that *CCDC170*, but not *ESR1*, mRNA levels were significantly higher in these normal (adjacent to tumor) tissues than in cancer tissue (Kruskal Wallis Test P <0.0001, (**Figure 2**). To investigate this unexpectedly high incidence in more detail, we analyzed normal breast tissues of ten women without diagnosed breast cancer, 16 benign fibroadenomas and 13 ductal carcinomas in situ (DCIS) tissues, all of them *ESR1*-positive. In addition, we measured the fusion transcripts in three sets of patient-matched normal breast and primary tumor carcinomas and four patient-matched sets of primary breast tumors and metastatic lymph nodes, also all *ESR1*-positive. In none of these cases did we detect an *ESR1* fusion transcripts with *AKAP12* or *ARMT1*. However, one of the breast tissues of women without breast cancer diagnosis (10%) showed *ESR1-CCDC170* exon 2 (E2-E2) fusion transcripts, one of the DCIS cases (7.7%) had *ESR1-CCDC170* exon 6 (E2-E6) fusion transcripts, and four patients with fibroadenoma (25%) had ESR1-*CCDC170* exon 8 (E2-E8) fusion transcripts (**Table 4** and **Table S3**). For one out of the three matched normal-tumor cases we found an *ESR1-CCDC170* exon 8 fusion in both the primary tumor and the normal breast tissue taken at a distance from the primary tumor. Finally, for two out of the four patients of which we had a matched primary tumor and lymph node metastasis, an *ESR1-CCDC170* exon 2 fusion was present in both the primary tumor and the lymph node metastasis.

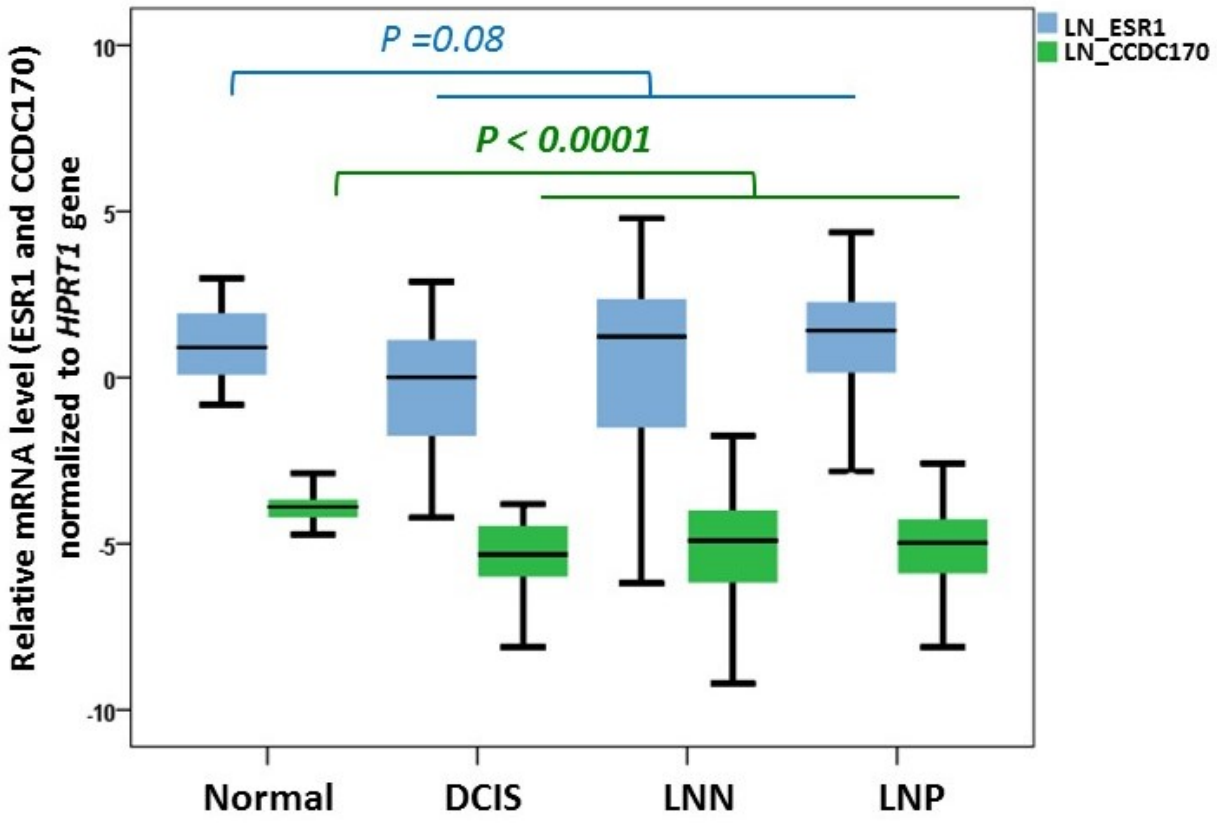


Figure 2: Expression of *ESR1* and *CCDC170* genes in breast tissues.

Relative *CCDC170* (blue box) and *ESR1* (green box) mRNA levels normalized to *HPRT1* gene levels are showed in the y-axis and were measured by RT-qPCR in normal (adjacent to tumor), benign (DCIS) and carcinoma (LNN and LNP) breast tissues. The box plots show interquartile ranges (IQR) together with the median (black horizontal line) of the *ESR1* and *CCDC170* mRNA levels for the different conditions. DCIS: ductal carcinomas in situ; LNP: Lymph node positive; LNN: Lymph node negative.

Prevalence of *ESR1* fusion genes in breast tumor tissues

Since fusion transcripts were predominantly detected in the *ESR1*+ population, we decided to investigate the clinical relevance of these transcripts in primary tumors. To this end, we stratified *ESR1+* patients in two distinct cohort: a predictive cohort of advanced BC patients treated with first-line endocrine therapy and a prognostic cohort of primary BC patients with lymph node negative disease (LNN) who did not receive any adjuvant systemic treatment.

In these two *ESR1*+ cohorts, *ESR1-ARMT1* fusion transcripts were detected in four patients of the predictive cohort (1.2%) and in three patients of the prognostic cohort (1.08%). Due to the low incidence of this *ESR1-ARMT1* fusion transcript, it was not further pursued. *ESR1-AKAP12* fusion transcripts were more common, and observed in 13 patients of the predictive cohort (4.04%) and in five patients of the prognostic cohort (1.79%). The *ESR1-CCDC170* fusion transcripts, however, were the most prevalent and detected in the predictive cohort in 89 patients (27.6%) and in the prognostic cohort in 66 patients (23.7%). Interestingly, all patients harboring an *ESR1-ARMT1* or an *ESR1-AKAP12* fusion were also positive for an *ESR1-CCDC170* rearranged transcript. Moreover, we noticed the coexistence of the three fusions in two subjects. Of all the732 breast tissue samples studied, the most prominent *ESR1-CCDC170* fusion transcripts found involved exon 2 of *ESR1* fused with exon 2 (18.93%) and exon 8 (16.86%) of *CCDC170* (**Table 4**).

Association of *ESR1* fusion genes with DFS and OS in the prognostic cohort

The presence of *ESR1-CCDC170* fusion transcripts in the primary tumor of our *ESR1+* LNN patients predicted a shorter disease-free survival in a Cox proportional hazards regression survival analysis (HR ± 95% CI: 1.44 (1.01 – 2.05), *P* = 0.044) **(Table 5).** We decided to investigate the two frequently present *ESR1-CCDC170* fusion transcripts (E2-E2 and E2-E8). Analyzing the *ESR1-CCDC170* exon 2 and exon 8 separately, showed that the fusion with exon 8 of *CCDC170* on its own associated with a short disease free survival (DFS; HR ± 95% CI: 1.95 (1.30 – 2.93), *P* = 0.001). No association with disease free survival was seen for *ESR1-AKAP12* fusion transcripts (HR ± 95% CI: 1.23 (0.39 – 3.87), *P* = 0.72). Concerning overall survival, only the presence of an *ESR1-CCDC170* exon 8 fusion predicted a shorter overall survival time (HR ± 95% CI: 1.85 (1.18 – 2.90, *P* = 0.007) The DFS and OS Kaplan Meier curves as a function of *ESR1-CCDC170* exon 8 fusion transcripts are shown in **Figure 3A** and **Figure 3B**, respectively. A multivariate analysis was performed in which age at primary surgery, pathological tumor classification, tumor grade, progesterone receptor and HER2 status were included. The analysis revealed HER2 status as a significant prognostic factor for overall survival, but not for DFS (*P* = 0.36) (**Table 5**). In these

analyses, the presence of *ESR1-CCDC170* exon 8 fusion transcripts was an independent prognostic factor for both DFS (HR ± 95% CI: 1.82 (1.20 – 2.75), *P* = 0.005) and OS (HR ± 95% CI: 1.71 (1.08 – 2.72), *P* = 0.001).

Table 5. Univariate and multivariate Cox proportional hazards regression survival analysis

		DFS				OS			
		Univariate analyses		Multivariate analyses		Univariate analyses		Multivariate analyses	
Parameters*	**n**	**HR (95% CI)**	***p-value***	**HR (95% CI)**	***p-value***	**HR (95% CI)**	***p-value***	**HR (95% CI)**	***p-value***
	279								
Age at primary surgery									
≤ 40 years	29	1.00		1.00		1.00		1.00	
41-50 years	81	0.59 (0.35-1.00)	***0.049***	0.60 (0.35-1.02)	*0.06*	0.53 (0.30-0.96)	***0.036***	0.51(0.28-0.94)	***0.032***
51-70 years	125	0.73 (0.44-1.19)	*0.20*	0.72 (0.44-1.18)	*0.19*	0.75 (0.44-1.28)	*0.30*	0.72 (0.42-1.26)	*0.25*
>70 years	44	0.78(0.43-1.40)	*0.41*	0.71 (0.39-1.28)	*0.25*	0.73 (0.37-1.43)	*0.35*	0.73 (0.37-1.47)	*0.38*
Menopausal status									
Premenopausal	120	1.00				1.00			
Postmenopausal	159	1.01 (0.73-1.38)	*0.96*			1.06 (0.74-1.53)	*0.73*		
Pathological Tumor classification									
pT1	151	1.00		1.00		1.00		1.00	
pT2 + unknown	119	1.54 (1.12-2.11)	***0.007***	1.35 (0.98-1.88)	0.069	1.30 (0.90-1.87)	*0.165*	1.19 (0.81-1.74)	*0.375*
pT3+pT4	9	2.31 (1.00-5.32)	***0.049***	2.47 (1.07-5.75)	**0.035**	3.26 (1.39-7.62)	***0.006***	3.45 (1.45-8.19)	***0.005***
Grade									
poor	131	1.00		1.00		1.00		1.00	
unknown	81	1.36 (0.97-1.91)	*0.076*	1.40 (0.98-1.99)	*0.064*	0.89(0.59-1.34)	*0.577*	0.97 (0.64-1.48)	*0.894*
moderate and good	67	0.52 (0.33-0.82)	***0.004***	0.57(0.36-.89)	***0.014***	0.51 (0.31-0.85)	***0.009***	0.57(0.34-0.94)	***0.029***
ER	279	1.11 (0.98-1.25)	*0.10*			0.99 (0.86-1.14)	*0.92*		
PR									
negative	62	1.00		1.00		1.00		1.00	
positive	217	0.66 (0.46-0.93)	***0.019***	0.68 (0.47-0.98)	***0.037***	0.49 (0.33-0.73)	***<0.001***	0.56 (0.37-0.85)	***0.007***
HER2 status*									
not amplified	239	1.00				1.00		1.00	
amplified	43	1.21 (0.80-1.84)	*0.36*			1.82 (1.17-2.84)	***0.008***	1.72 (1.08-2.73)	***0.022***
Predictive cohort									
1st line Tamoxifen	235								
at least one *ESR1-CCDC170* (exon 2 to 8) fusion		0.96 (0.71-1.30)	*0.81*			1.16 (0.85-1.60)	0.35		

ESR1-AKAP12		1.37 (0.61-3.10)	*0.44*			1.92 (0.84-4.35)	0.12		
1st line AI	87								
at least one *ESR1-CCDC170* (exon 2 to 8) fusion		0.85 (0.53-1.37)	*0.50*						
ESR1-AKAP12		1.62 (0.73-3.60)	*0.24*						

		Separately added to the base analyses							
		DFS				**OS**			
		Univariate analyses		**Multivariate analyses**		**Univariate analyses**		**Multivariate analyses**	
Parameters ESR1+ LNN	**n**	**HR (95% CI)**	***p-value***	**HR (95% CI)**	***p-value***	**HR (95% CI)**	***p-value***	**HR (95% CI)**	***p-value***
at least one ESR1-CCDC170 (exon 2 to 8) fusion									
negative	213	1.00		1.00		1.00		1.00	
positive	66	1.44 (1.01-2.05)	***0.044***	1.33 (0.92-1.92)	*0.13*	1.67 (1.13-2.47)	***0.010***	1.54(1.02-2.33)	***0.042***
ESR1-CCDC170 (exon 2) fusion									
negative	246	1.00				1.00		1.00	
positive	33	1.40 (0.89-2.21)	*0.14*			1.75(1.07-2.87)	***0.026***	1.38 (0.82-2.33)	*0.22*
ESR1-CCDC170 (exon 8) fusion									
negative	240	1.00		1.00		1.00		1.00	
positive	39	1.95 (1.30-2.93)	***0.001***	1.82 (1.20-2.75)	***0.005***	1.82 (1.18-2.90)	***0.007***	1.71(1.08-2.72)	***0.001***
ESR1-AKAP12									
negative	274	1.00				1.00			
positive	5	1.23 (0.39-3.87)	*0.72*			2.45 (0.90-6.65)	*0.08*		

n: number of patients; DFS: disease free survival; OS: overall survival; HR: hazard ratio; CI: confidence interval; pT1: small primary tumor (tumour is 2 centimetre across or less); pT2: tumour more than 2 centimetres but no more than 5 centimetres across; pT3: T3 tumour bigger than 5 centimetres across; pT4: tumor with phatological stage; ER: estrogen receptor; PR: PR: progesterone receptor; HER2: human epidermal growth factor receptor; AI: aromatase inhibitors; *ESR1-CCDC170*: *ESR1-CCDC170* gene fusion; *ESR1-AKAP12*: *ESR1-AKAP12* gene fusion. Statistically significant differences are indicated in bold. * due to unknown data numbers do not add up to 279.

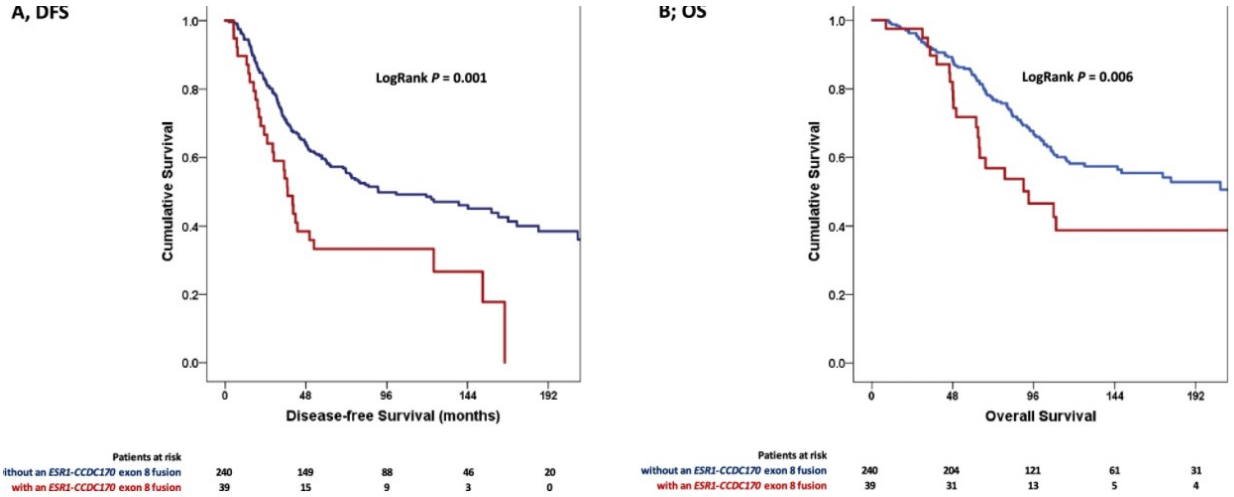


Figure 3: Disease free survival (DFS) and overall survival (OS) in the prognostic cohort.

The DFS and OS Kaplan Meier curves in positive LNN patients. A) DFS of patients with or without *ESR1-CCDC170* exon 8 fusion gene; (B) OS of patients with or without *ESR1-CCDC170* exon 8 fusion gene. The reported *P*-value is from a log-rank test and the test statistics from Cox regression analyses.

Association of *ESR1* fusion genes with clinical characteristics, PFS and post-relapse overall survival in advanced BC patients

The fusion transcripts were related with traditional clinical parameters, with response to first-line endocrine therapy in the predictive cohort (n=322; tamoxifen (n=235), aromatase inhibitors (n=87)) (**Table 2)**. In the predictive cohort *ESR1-CCDC170* fusion transcripts showed an association with age at start of first-line treatment, whereas *ESR1-AKAP12* fusion transcripts were enriched in patients with progesterone-negative primary tumors at time of surgery and in AI-treated patients who received adjuvant tamoxifen. No relation with PFS after first-line tamoxifen (n=235) was found in our Cox proportional hazards regression survival analysis for the *ESR1-CCDC170* fusion transcripts (HR ± 95% CI: 0.96 (0.71 – 1.30), P = 0.81) nor for the *ESR1-AKAP12* fusion transcripts (HR ± 95% CI: 1.37 (0.61 – 3.10), P = 0.44) **(Table 5)**. In addition, the presence of these fusion transcripts did not affect the time from relapse to death (post-relapse survival, HR ± 95% CI: 1.16 (0.85 – 1.60), P = 0.35 and 1.92 (0.84 – 4.35), P = 0.12, for *ESR1* fusions with *CCDC170* and *AKAP12*, respectively) **(Table 5)**. Similarly, also no association with PFS for first-line aromatase inhibitors (n=87) was found for *ESR1-CCDC170* fusion transcripts (HR ± 95% CI: 0.85 (0.53 – 1.37), P = 0.50) nor for the *ESR1-AKAP12* fusion transcripts (HR ± 95% CI: 1.62

(0.73 – 3.60), P = 0.24). With data available for only 27 patients post-relapse, we did not analyze post-relapse survival for aromatase inhibitors. Moreover, no-significant associations with PFS were seen when the *ESR1-CCDC170* exon 2 and exon 8 fusion transcripts were analyzed separately **(Table 5)**.

Discussion

The genetic landscape contributing to de novo or acquired resistance to endocrine therapy in breast cancer patients is not completely understood yet. In this study, we investigated the occurrence of recurrent fusion transcripts between *ESR1* and three different loci adjacent to *ESR1* (*CCDC170*, *AKAP12* and *ARMT1*) and correlated their presence with clinical outcome. All of the fusion transcripts analyzed are recurrent and most frequently present in ER-positive disease and among them *ESR1-CCDC170* fusion transcripts were the most predominant. As proposed by others [10, 13], the presumption was that these fusion transcripts, which are considered to cause constitutive ER signaling, might signify resistance to endocrine therapy. However, in patients with advanced breast cancer, we did not find that the presence of any of these fusion transcripts is associated with outcome to endocrine therapy whether it concerned first line tamoxifen or an aromatase inhibitor. Importantly, smaller size effects from these the variants may be undetected due to the relatively small sample size of the study cohort, 87 patients treated with aromatase inhibitors and 235 subjects with tamoxifen. In contrast, in patients with primary BC and not receiving adjuvant systemic hormone treatments, we found that fusion between *ESR1* and *CCDC170* in general, and between exon 2 of *ESR1* and exon 8 of *CCDC170* in particular, predicted in uni- and multivariable analyses shorter disease free survival as well as shorter overall survival. Thus, *ESR1* and *CCDC170* fusion transcript pinpoint cancers with an adverse outcome.

Understanding the molecular mechanisms that underlay the origin of fusion transcripts could help to comprehend the role of these fusions in carcinogenesis as well as improve the diagnosis of cancer patients [10, 13]. Although the progress in DNA sequencing enhanced detection of recurrent and pathological breast cancer fusions, the complexity of underlying genomic rearrangement patterns makes their characterization at the DNA level often difficult. The fusion between *ESR1* and its neighboring gene *CCDC170* are potentially generated by tandem duplication [9, 13, 27, 28], which is also causing other genetic rearrangements in cancer [9, 29, 30]. Kim et al. found a region within the *ESR1* genomic locus most vulnerable to DNA strand

breakage, which often included intron 6 region of its neighboring gene *CCDC170*, resulting in oncogenic mRNA *ESR1-CCDC170* fusion transcript of exons 2 of *ESR1* connected to exon 2-11 of *CCDC170*, i.e. the C-terminal domain of *CCDC170* [31]. Irrespective of mechanisms causing the gene fusions, they occur in a patient-specific manner, which makes their identification at the DNA level less suitable for routine diagnostics. Our method to analyze fusion transcripts is much less dependent on exact position of the underlying gene fusion at the DNA level and is therefore better suited to evaluate as a general biomarker in large patient cohorts. However, an important caveat for detecting gene fusions at the transcript level is the fact that it cannot distinguish between fusion transcripts arising from actual genetic rearrangements and those that arise from transcription reading from one gene into the next without a genetic cause. Interestingly, Giltnane et al rejected the option of a run-on transcription for these genes since the 5'end of *ESR1* is fused to the 3'ends of *CCDC170* and *AKAP12*, which are upstream of ESR1 gene [10]. Finally, the generation of artefactual fusion sequences, which are randomly ligated during the sequencing procedure, might happen, as previously reported by Veeraraghavan et al. [13]. Overall, we performed RT-qPCR analysis and investigated RNA not DNA, therefore we cannot tell whether fusion transcripts are the results of (DNA) rearrangements. Furthermore, to our great surprise, *ESR1-CCDC170* and *ESR1-AKAP12* fusions were detected in ER-negative patients even if at low frequency (2.3% and 0.76%, respectively). Besides sampling bias, this finding might be explained by a challenge in ER and PR determination. Althought immunohistochemistry (IHC) is the "gold standard" to determine the surrogate markers ER and PR for breast cancer classification, several studies addressed limitations in IHC by shedding light on the discordance rates in scoring hormone receptor status with negative and false-positive rates in ER and PR statuses higher than 20% [32, 33]. Similarly, a recently article by Fakhri et all. found that 12.5% of samples negative for ER by IHC were positive via microarray analysis [34]. In this context, we performed RT-qPCR to accurately determine hormone receptor status. However, this method could be subject to bias during RNA measurement. Moreover, a recently study found that in primary breast cancers, the ER-negative phenotype is not the result of mutations in ER gene, but is due to deficient ER expression at the transcriptional or post-transcriptional level [35]. Therefore, we might hypothesize that the ER expression might be restored in ER-negative patients due to the strongly impact of the signaling environment, as already demonstrated for breast cancer cells via inhibition of DNA methylation or histone deacetylation [36].

Another interesting question regards the biological significance of clinically relevant fusion transcripts. Gene fusions and their products (RNAs and proteins) are assumed to be exclusive to cancer. However, RNA-sequencing analyses from normal appearing margins of cancerous specimens showed fusion transcripts also in normal tissues [37]. In fact, oncogenic rearrangements, such as the *EML4-ALK* [38], *NPM-ALK* [39], *JAZF1-JJAZ1* [40] and *BCR-ABL1* [41] fusions are also expressed at a low level in histologically non-neoplastic tissues [9]. In our study, expression of *ESR1* fused to exons 2 and exon 8 of *CCDC170* was found in mammary epithelial tissues derived from women without diagnosis of breast cancer, and in cases with (benign) fibroadenomas, respectively. Also in early stages of breast cancer, like DCIS, we detected fusion transcripts. Moreover, *ESR1-CCDC170* fusion transcripts were also detected in normal breast tissues of patients with diagnosed breast cancer. This argues that a percentage may be transcript read-through instead of fusion transcripts arising from gene fusions.

According to our results, the expression of *ESR1-CCDC170* exon 2 and exon 8 fusion transcripts were linked to a less favorable disease in BC patients who not received adjuvant systemic treatment. Overall, our results are in agreement with those reported by Veeraraghavan et al. which showed that *ESR1-CCDC170* fusions, when introduced into ER-positive breast cancer cells, leads to a markedly increase of cell motility and colony-forming ability, increase in S-G2/M phase cells and a decrease in G0/G1 phase cells. Although several functional studies [9, 42] demonstrated a role of *ESR1-CCDC170* fusions in endocrine therapy resistance, no relationship between fusion transcripts and treatment outcome was observed in our predictive cohort. Overall, since *ESR1-CCDC170* fusions in our study demonstrated no predictive value for endocrine therapy resistance, their prognostic value might be explained by the recurrent incidence of read-through events during cell cycle progression. This latter has been exemplified with the abundance of *CTSD-IFITM10* readthrough fusions during breast cancer cell proliferation [43].

Conclusions

The most important conclusion from our work is that among the fusion transcripts evaluated measuring *ESR1-CCDC170* exon 8 fusion transcripts in primary breast cancers has diagnostic potential as it identifies a more aggressive subset of ER-positive breast cancer patients. Furthermore, with our study we demonstrated that *ESR1-CCDC170* fusion transcript does not predict endocrine therapy resistance in our setting.

References

1. Pritchard, K.I., *Endocrine therapy: is the first generation of targeted drugs the last?* J Intern Med, 2013. **274**(2): p. 144-52.

2. Riggins, R.B., et al., *Pathways to tamoxifen resistance.* Cancer Lett, 2007. **256**(1): p. 1-24.

3. Osborne, C.K. and R. Schiff, *Mechanisms of endocrine resistance in breast cancer.* Annu Rev Med, 2011. **62**: p. 233-47.

4. Schiavon, G., et al., *Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer.* Sci Transl Med, 2015. **7**(313): p. 313ra182.

5. Pejerrey, S.M., et al., *The Impact of ESR1 Mutations on the Treatment of Metastatic Breast Cancer.* Horm Cancer, 2018. **9**(4): p. 215-228.

6. Beije, N., et al., *Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients.* Mol Oncol, 2018. **12**(1): p. 48-57.

7. Vitale, S.R., et al., *An Optimized Workflow to Evaluate Estrogen Receptor Gene Mutations in Small Amounts of Cell-Free DNA.* J Mol Diagn, 2019. **21**(1): p. 123-137.

8. Robinson, D.R., et al., *Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer.* Nat Med, 2011. **17**(12): p. 1646-51.

9. Veeraraghavan, J., et al., *Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers.* Nat Commun, 2014. **5**: p. 4577.

10. Giltnane, J.M., et al., *Genomic profiling of ER(+) breast cancers after short-term estrogen suppression reveals alterations associated with endocrine resistance.* Sci Transl Med, 2017. **9**(402).

11. Li, S., et al., *Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts.* Cell Rep, 2013. **4**(6): p. 1116-30.

12. Yoshihara, K., et al., *The landscape and therapeutic relevance of cancer-associated transcript fusions.* Oncogene, 2015. **34**(37): p. 4845-54.

13. Veeraraghavan, J., et al., *Recurrent and pathological gene fusions in breast cancer: current advances in genomic discovery and clinical implications.* Breast Cancer Res Treat, 2016. **158**(2): p. 219-32.

14. Sieuwerts, A.M., et al., *How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study.* Clin Cancer Res, 2005. **11**(20): p. 7311-21.

15. Sieuwerts, A.M., et al., *Concentrations of TIMP1 mRNA splice variants and TIMP-1 protein are differentially associated with prognosis in primary breast cancer.* Clin Chem, 2007. **53**(7): p. 1280-8.

16. Ramirez-Ardila, D.E., et al., *Hotspot mutations in PIK3CA associate with first-line treatment outcome for aromatase inhibitors but not for tamoxifen.* Breast Cancer Res Treat, 2013. **139**(1): p. 39-49.

17. Hayward, J.L., et al., *Assessment of response to therapy in advanced breast cancer. A project of the programme on clinical oncology of the International Union against Cancer, Geneva, Switzerland.* Eur J Cancer, 1978. **14**(11): p. 1291-2.

18. Onstenk, W., et al., *Efficacy of Cabazitaxel in Castration-resistant Prostate Cancer Is Independent of the Presence of AR-V7 in Circulating Tumor Cells.* Eur Urol, 2015. **68**(6): p. 939-45.

19. Sieuwerts, A.M., et al., *Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR.* Breast Cancer Res Treat, 2009. **118**(3): p. 455-68.

20. Sieuwerts, A.M., et al., *mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients.* Clin Cancer Res, 2011. **17**(11): p. 3600-18.

21. Lei, J.T., X. Gou, and M.J. Ellis, *ESR1 fusions drive endocrine therapy resistance and metastasis in breast cancer.* Mol Cell Oncol, 2018. **5**(6): p. e1526005.

22. Paratala, B.S., et al., *Emerging Role of Genomic Rearrangements in Breast Cancer: Applying Knowledge from Other Cancers.* Biomark Cancer, 2016. **8**(Supple 1): p. 1-14.

23. Jones, D.T., et al., *Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas.* Cancer Res, 2008. **68**(21): p. 8673-7.

24. Lipson, D., et al., *Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies.* Nat Med, 2012. **18**(3): p. 382-4.

25. Kim, R.N., et al., *Perspective Insight into Future Potential Fusion Gene Transcript Biomarker Candidates in Breast Cancer.* Int J Mol Sci, 2018. **19**(2).

26. Babiceanu, M., et al., *Recurrent chimeric fusion RNAs in non-cancer tissues and cells.* Nucleic Acids Res, 2016. **44**(6): p. 2859-72.

27. Martelli, M.P., et al., *EML4-ALK rearrangement in non-small cell lung cancer and non-tumor lung tissues.* Am J Pathol, 2009. **174**(2): p. 661-70.

28. Maes, B., et al., *The NPM-ALK and the ATIC-ALK fusion genes can be detected in non-neoplastic cells.* Am J Pathol, 2001. **158**(6): p. 2185-93.

29. Li, H., et al., *A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells.* Science, 2008. **321**(5894): p. 1357-61.

30. Biernaux, C., et al., *Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals.* Blood, 1995. **86**(8): p. 3118-22.

31. Li Li, J.V., Yiheng Hu, Xian Wang, Ying Tan, Rachel Schiff and Xiaosong Wang, *Abstract 376: Therapeutic role of ESR1-CCDC170 gene fusion in breast cancer endocrine resistance.* American Association for Cancer Research, 2019. **79**(13): p. Supplement, pp. 376.

32. Varley, K.E., et al., *Recurrent read-through fusion transcripts in breast cancer.* Breast Cancer Res Treat, 2014. **146**(2): p. 287-97.

Supplementary Information

Table S1. Assays used for *ESR1-CCDC170* fusion analyses

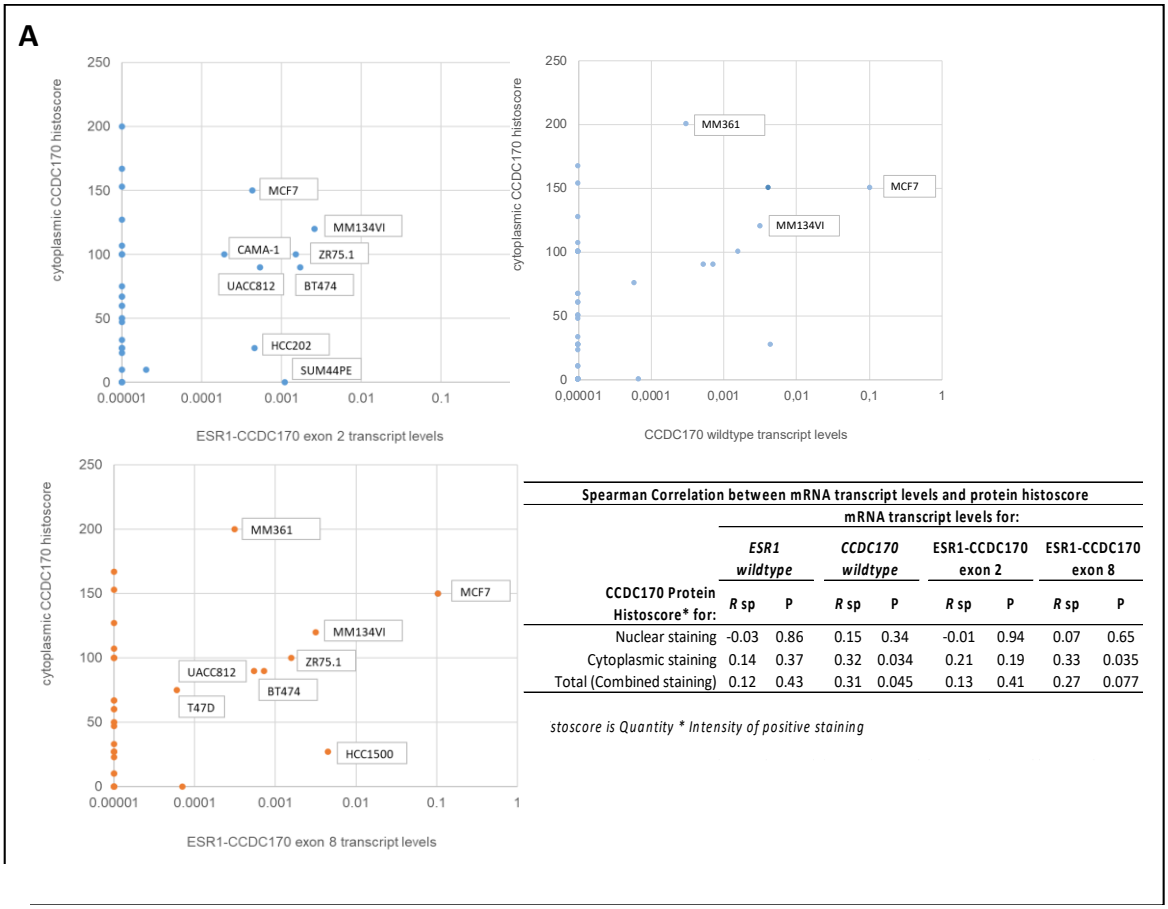
Primer	Type	Sequence(5'->3')	Exon size (bp)
ESR1_F2_A	F	CTGCGGTACCAAATATCAGCAC	146
CCDC170_R2	R	TTTTGAGCCACATTCCGATAG	129
CCDC170_R3	R	AGCTTTCAGTTCTTGACAGGAGA	257
CCDC170_R4	R	CAAGCAGTCACGCAGTTGAG	145
CCDC170_R5	R	TGCTTCCATCTCATGGACATT	186
CCDC170_R6	R	TCCTGACTGGCCTTCAAACTC	318
CCDC170_R7	R	AACCCAGACTCCTTTCCCAAC	201
CCDC170_R8	R	CAGCCATCTGGTCCAACTTC	174
CCDC170_R10	R	CTTCTCCAGTTGGTCTCTGGAT	243
CCDC170_R11	R	AAGCCTAGCATCTGCGACAC	237

F: forward primer; R: reverse primer

Table S2. Expected *ESR1-CCDC170* fusion products

Forward primer	Reverse primer	product size fusion/exon	*R2*	*R3*	*R4*	*R5*	*R6*	*R7*	*R8*	*R9*	*R10*
ESR1_F2	CCDC170_R2	144	144								
	CCDC170_R3	124	253	124							
	CCDC170_R4	164	550	421	164						
	CCDC170_R5	158	689	560	303	158					
	CCDC170_R6	238	955	826	569	424	238				
	CCDC170_R7	135	1170	1041	784	639	453	135			
	CCDC170_R8	125	1361	1232	975	830	644	326	125		
	CCDC170_R9	150	1560	1431	1174	1029	843	525	324	150	
	CCDC170_R10	131	1789	1660	1403	1258	1072	754	553	379	136

In blue are showed the expected size of fusion products while in violet are depicted the aspecific products which can be generated durring qRT-PCR



Spearman Correlation between mRNA transcript levels and protein histoscore								
	mRNA transcript levels for:							
	ESR1 wildtype		*CCDC170 wildtype*		ESR1-CCDC170 exon 2		ESR1-CCDC170 exon 8	
CCDC170 Protein Histoscore* for:	*R* sp	P	*R* sp	P	*R* sp	P	*R* sp	P
Nuclear staining	-0.03	0.86	0.15	0.34	-0.01	0.94	0.07	0.65
Cytoplasmic staining	0.14	0.37	0.32	0.034	0.21	0.19	0.33	0.035
Total (Combined staining)	0.12	0.43	0.31	0.045	0.13	0.41	0.27	0.077

*stoscore is Quantity * Intensity of positive staining*

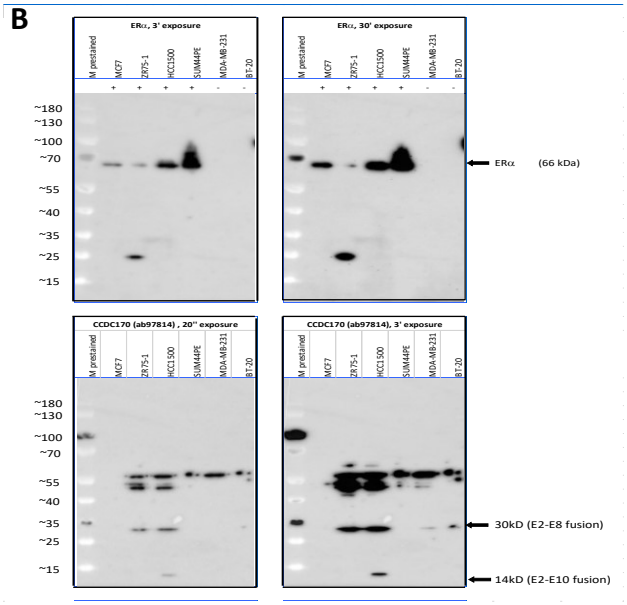


Figure S1. Expression of CCDC170 wildtype and fusion protein evaluated by immunohistochemical staining (IHC) and western blotting in breast cancer cell lines.

A. IHC performed on a cell line microarray of 44 breast cancer cell lines show a histoscore correlation between the cytoplasmic CCDC170 and CCDC170 wildtype as well as between ESR1-CCDC170 exon 8 fusion transcript levels and CCDC170 wildtype, but not with exon 2 fusion transcript levels. B. Western blotting analyses demonstrated the expression of CCDC170 fusion protein. The exon 2 ESR1 – exon 8

CCDC170 fusion product (~30kD) was detected in ZR75.1 and HCC1500, but not in MCF-7. The exon 2-exon 10 CCDC170 fusion protein (~14kD) was also observed, but only in HCC1500.

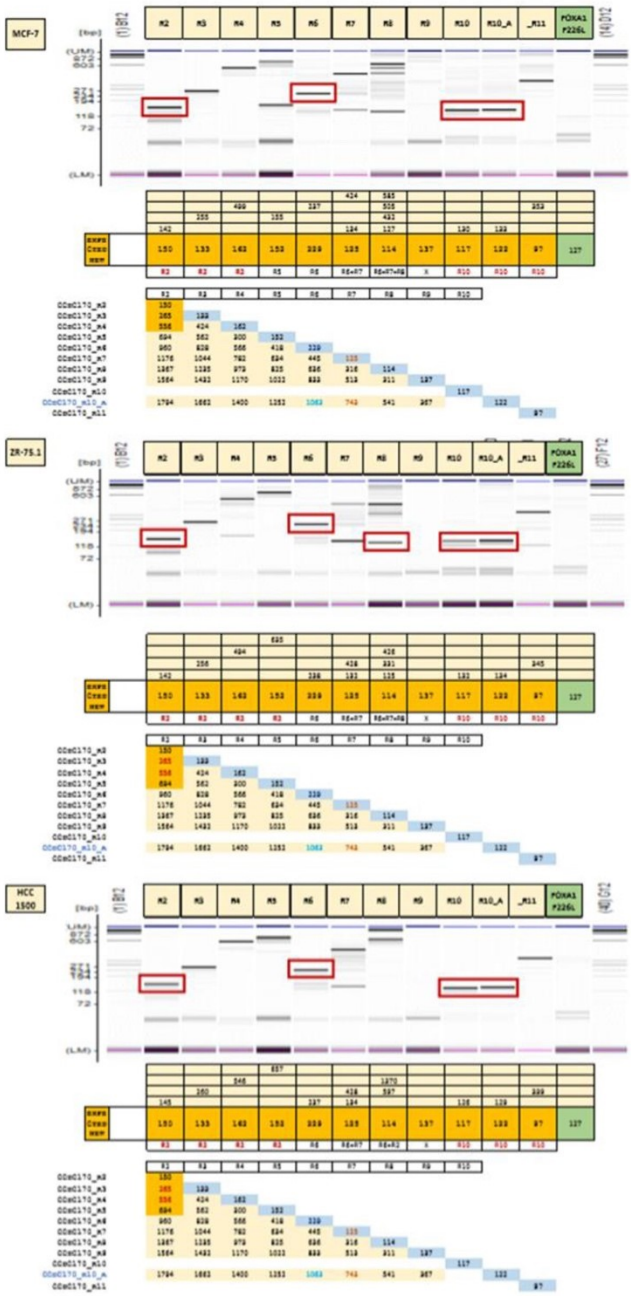


Figure S2. *ESR1-CCDC170* fusions confirmation by MultiNA in a subset of Breast cancer cell lines.

If the Taqman probe-based RT-qPCR generated a positive Cq value, the expected fusion gene product sizes were validated by MultiNA. Only products with a MultiNA fusion product of the expected size and a ΔCq ≥ -25 relative to the two reference genes were considered positive for the fusion product. MultiNA analyses confirmed the *CCDC170* RNA fusion products in three breast cancer cell lines. Red boxes indicate fusion expression with correct fragment sizes.

Figure S3. Expression of *ESR1* fusions and reference *(ESR1* and *CCDC170)* genes in a panel of 54 breast cancer cell lines.

Cell line name	*ESR1-CCDC170 exon 2*	*ESR1-CCDC170 exon 3*	*ESR1-CCDC170 exon 4*	*ESR1-CCDC170 exon 6*	*ESR1-CCDC170 exon 7*	*ESR1-CCDC170 exon 8*	*ESR1-CCDC170 exon 10*	*ESR1-AKAP12*	*ESR1-ARMT1*	*ESR1*	*CCDC170*
SUM-44PE	0,00109	0,00091	0,00001	0,00069	0,00028	0,00007	0,00001	0,000001	0,000001	2,06360	0,02287
BT474	0,00170	0,00129	0,00001	0,00170	0,00247	0,00072	0,00004	0,000001	0,000001	0,47363	0,01481
HCC 1419	0,00006	0,00001	0,00001	0,00002	0,00001	0,00001	0,00001	0,000001	0,000001	0,70637	0,00995
MDA-MB-361	0,00001	0,00031	0,00001	0,00021	0,00088	0,00031	0,00003	0,000001	0,000001	0,90239	0,00812
UACC-812	0,00054	0,00001	0,00013	0,00021	0,00001	0,00054	0,00001	0,000001	0,000001	0,44705	0,00699
BT483	0,00001	0,00001	0,00001	0,00018	0,00001	0,00001	0,00001	0,000001	0,000001	1,61907	0,00466
MDA-MB-175VII	0,00001	0,00001	0,00001	0,00008	0,00001	0,00001	0,00001	0,000001	0,000001	0,42196	0,00455
MDA-MB-134VI	0,00258	0,00110	0,00103	0,00121	0,00035	0,00314	0,00031	0,000001	0,000001	4,06098	0,00438
MCF-7	0,00043	0,00047	0,00045	0,42730	0,36687	0,10415	0,00811	0,000001	0,000001	1,90769	0,00420
T47D	0,00001	0,00001	0,00001	0,00001	0,00001	0,00006	0,00001	0,000001	0,000001	1,36777	0,00416
HCC 1500	0,00001	0,00001	0,00030	0,00912	0,01026	0,00448	0,01599	0,000001	0,000001	2,34322	0,00323
CAMA-1	0,00019	0,00013	0,00018	0,00021	0,00023	0,00001	0,00001	0,000063	0,000008	0,65603	0,00280
MDA-MB-330	0,00005	0,00001	0,00001	0,00002	0,00003	0,00001	0,00001	0,000001	0,000001	0,43684	0,00223
MPE600	0,00001	0,00001	0,00001	0,00005	0,00001	0,00001	0,00001	0,000001	0,000001	1,06081	0,00187
ZR-75.1	0,00151	0,00088	0,00149	0,00250	0,00175	0,00158	0,00016	0,000003	0,000001	1,19900	0,00162
HCC 2218	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00611	0,00088
OCUB-M	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00485	0,00044
SUM-1315MO2	0,00001	0,00001	0,00001	0,00004	0,00001	0,00001	0,00001	0,000001	0,000001	0,00017	0,00020
ZR-75.30	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,66672	0,00019
MDA-MB-436	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00006	0,00008
HCC 1569	0,00001	0,00001	0,00001	0,00011	0,00002	0,00001	0,00001	0,000001	0,000001	0,02084	0,00005
MDA-MB-415	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	1,24415	0,00004
HCC 1599	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,06172	0,00003
MDA-MB-157	0,00001	0,00001	0,00001	0,00001	0,00002	0,00001	0,00001	0,000001	0,000001	0,00009	0,00003

Cell line name	*ESR1-CCDC170* exon 2	*ESR1-CCDC170* exon 3	*ESR1-CCDC170* exon 4	*ESR1-CCDC170* exon 6	*ESR1-CCDC170* exon 7	*ESR1-CCDC170* exon 8	*ESR1-CCDC170* exon 10	*ESR1-AKAP12*	*ESR1-ARMT1*	*ESR1*	*CCDC170*
HCC 202	0,00046	0,00001	0,00001	0,00003	0,00001	0,00001	0,00001	0,000001	0,000001	0,00605	0,00003
BT549	0,00001	0,00001	0,00001	0,00001	0,00002	0,00001	0,00001	0,000001	0,000001	0,00001	0,00003
OCUB-F	0,00001	0,00001	0,00001	0,00002	0,00001	0,00001	0,00001	0,000001	0,000001	0,00109	0,00002
HCC 38	0,00001	0,00001	0,00001	0,00002	0,00001	0,00001	0,00001	0,000001	0,000001	0,00001	0,00002
HCC 1395	0,00001	0,00001	0,00001	0,00002	0,00001	0,00001	0,00001	0,000001	0,000001	0,00001	0,00001
MDA-MB-231	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00004	0,00001
SK-BR-7	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00006	0,00001
EVSA-T	0,00002	0,00001	0,00001	0,00003	0,00001	0,00001	0,00001	0,000001	0,000001	0,00007	0,00001
SUM-159PT	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00018	0,00001
Hs578T	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00021	0,00001
MDA-MB-435s	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00024	0,00001
MDA-MB-453	0,00001	0,00001	0,00001	0,00003	0,00001	0,00001	0,00001	0,000001	0,000001	0,00054	0,00001
SK-BR-3	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00428	0,00001
HCC 1187	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00428	0,00001
SUM-102PT	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00473	0,00001
SUM-225CWN	0,00001	0,00001	0,00001	0,00004	0,00001	0,00001	0,00001	0,000001	0,000001	0,00939	0,00001
SUM-190PT	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,00976	0,00001
HCC 70	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,01042	0,00001
UACC-893	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,02093	0,00001
MDA-MB-468	0,00001	0,00001	0,00001	0,00003	0,00001	0,00001	0,00001	0,000001	0,000001	0,02137	0,00001
SUM-229 PE	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,02693	0,00001
SUM-149PT	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,02693	0,00001
BT20	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,02988	0,00001
HCC 1954	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,03377	0,00001
HCC 1143	0,00001	0,00001	0,00001	0,00002	0,00001	0,00001	0,00001	0,000001	0,000001	0,03561	0,00001
HCC 1806	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,04539	0,00001
SK-BR-5	0,00004	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,07408	0,00001

Cell line name	*ESR1-CCDC170 exon 2*	*ESR1-CCDC170 exon 3*	*ESR1-CCDC170 exon 4*	*ESR1-CCDC170 exon 6*	*ESR1-CCDC170 exon 7*	*ESR1-CCDC170 exon 8*	*ESR1-CCDC170 exon 10*	*ESR1-AKAP12*	*ESR1-ARMT1*	*ESR1*	*CCDC170*
SUM-185PE	0,00001	0,00001	0,00001	0,00002	0,00001	0,00001	0,00001	0,000001	0,000001	0,07921	0,00001
HCC-1937	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,09551	0,00001
SUM-52PE	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,000001	0,000001	0,50063	0,00001

Genes with expression level above the threshold are indicated in orange.

Table S3. Details of prevalence of *ESR1-CCDC170* (exon 1-11) fusion genes in the different analyzed cohorts.

Part A			ESR1-CCDC170													
			exon 1			exon 2			exon 3			exon 4			exon 5		
Total Count			**no**	**yes**		**no**	**yes**		**no**	**yes**		**no**	**yes**		**no**	**yes**	
732*	Patients studied	ESR1-	131	0	**0,00%**	128	3	**2,29%**	131	0	**0,00%**	131	0	**0,00%**	131	0	**0,00%**
		ESR1 +	597	4	**0,5%**	513	88	**14.64%**	596	5	**0,70%**	596	5	**0,70%**	599	2	**0,27%**
235	1st line Tamoxifen	ESR1-															
		ESR1 +	234	1	**0,43%**	204	31	**13,19%**	234	1	**0,43%**	233	2	**0,85%**	235	0	**0,00%**
87	1st line AI	ESR1-															
		ESR1 +	85	2	**2,30%**	68	19	**21,84%**	85	2	**2,30%**	85	2	**2,30%**	86	1	**1,15%**
322	Predictive cohort	ESR1-															
		ESR1 +	319	3	**0,93%**	272	50	**15,53%**	319	3	**0,93%**	318	4	**1,24%**	321	1	**0,31%**
369	Primary LNN cohort investigated	ESR1-	90	0	**0,00%**	87	3	**3,33%**	90	0	**0,00%**	90	0	**0,00%**	90	0	**0,00%**
		ESR1 +**	278	1	**0,36%**	246	33	**11,83%**	277	2	**0,72%**	279	0	**0,00%**	278	1	**0,36%**
36	Normal breast tissue of breast cancer patients	ESR1-	0	0		0	0		0	0		0	0		0	0	
		ESR1 +	35	1	**2,8%**	18	18	**50,0%**	35	1	**2,8%**	36	0	**0,0%**	36	0	**0,0%**
16	Tissue of breast fibroadenoma's	ESR1-	0	0		0	0		0	0		0	0		0	0	
		ESR1 +	16	0	**0,0%**	16	0	**0,0%**	16	0	**0,0%**	16	0	**0,0%**	16	0	**0,0%**
13	Tissue of breast DCIS	ESR1-															
		ESR1 +	13	0	**0,0%**	13	0	**0,0%**	13	0	**0,0%**	13	0	**0,0%**	13	0	**0,0%**
10	Normal breast tissue of healthy women	ESR1-	0	0		0	0		0	0		0	0		0	0	
		ESR1 +	10	0	**0,0%**	9	1	**10,0%**	10	0	**0,0%**	10	0	**0,0%**	10	0	**0,0%**

ESR1: Estrogen Receptor 1 gene; AI: Aromatase Inhibitor; LNP: Lymph node positive; LNN: Lymph node negative; DCIS: ductal carcinomas in situ; .* not all patients were further analyzed; **: included into the study.

Part B			ESR1-CCDC170													
			exon 6			exon 7			exon 8			exon 10			exon 11		
Total Count			**no**	**yes**		**no**	**yes**		**no**	**yes**		**no**	**yes**		**no**	**yes**	
732*	Patients studied	ESR1-	131	0	**0,00%**	130	1	**0,76%**	130	1	**0,76%**	131	0	**0,00%**	131	0	**0,00%**
		ESR1 +	581	20	**2,70%**	581	20	**2,70%**	504	97	**16,1%**	591	35	**4,8%**	596	5	**0,70%**
235	1st line Tamoxifen	ESR1-															
		ESR1 +	230	5	**2,13%**	226	9	**3,83%**	201	34	**14,47%**	222	13	**5,53%**	233	2	**0,85%**
87	1st line AI	ESR1-															
		ESR1 +	80	7	**8,05%**	85	2	**2,30%**	70	17	**19,54%**	81	6	**6,90%**	87	0	**0,00%**
322	Predictive cohort	ESR1-															
		ESR1 +	310	12	**3,73%**	311	11	**3,42%**	271	51	**15,84%**	303	19	**5,90%**	320	2	**0,62%**
369	Primary LNN cohort investigated	ESR1-	90	0	**0,00%**	89	1	**1,11%**	89	1	**1,11%**	90	0	**0,00%**	90	0	**0,00%**
		ESR1 +**	270	9	**3,23%**	268	11	**3,94%**	240	39	**13,98%**	268	11	**3,94%**	276	3	**1,08%**
36	Normal breast tissue of breast cancer patients	ESR1-	0	0		0	0		0	0		0	0		0	0	
		ESR1 +	35	1	**2,8%**	35	1	**2,8%**	23	13	**36,1%**	34	2	**5,6%**	36	0	**0,0%**
16	Tissue of breast fibroadenoma's	ESR1-	0	0		0	0					0	0		0	0	
		ESR1 +	16	0	**0,0%**	16	0	**0,0%**	12	4	**25,0%**	16	0	**0,0%**	16	0	**0,0%**
13	Tissue of breast DCIS	ESR1-															
		ESR1 +	12	1	**7,7%**	13	0	**0,0%**	13	0	**0,0%**	12	1	**7,7%**	13	0	**0,0%**
10	Normal breast tissue of healthy women	ESR1-	0	0		0	0		0	0		0	0		0	0	
		ESR1 +	10	0	**0,0%**	10	0	**0,0%**	10	0	**0,0%**	10	0	**0,0%**	10	0	**0,0%**

ESR1: Estrogen Receptor 1 gene; AI: Aromatase Inhibitor; LNP: Lymph node positive; LNN: Lymph node negative; DCIS: ductal carcinomas in situ.* not all patients were further analyzed; **: included into the study.

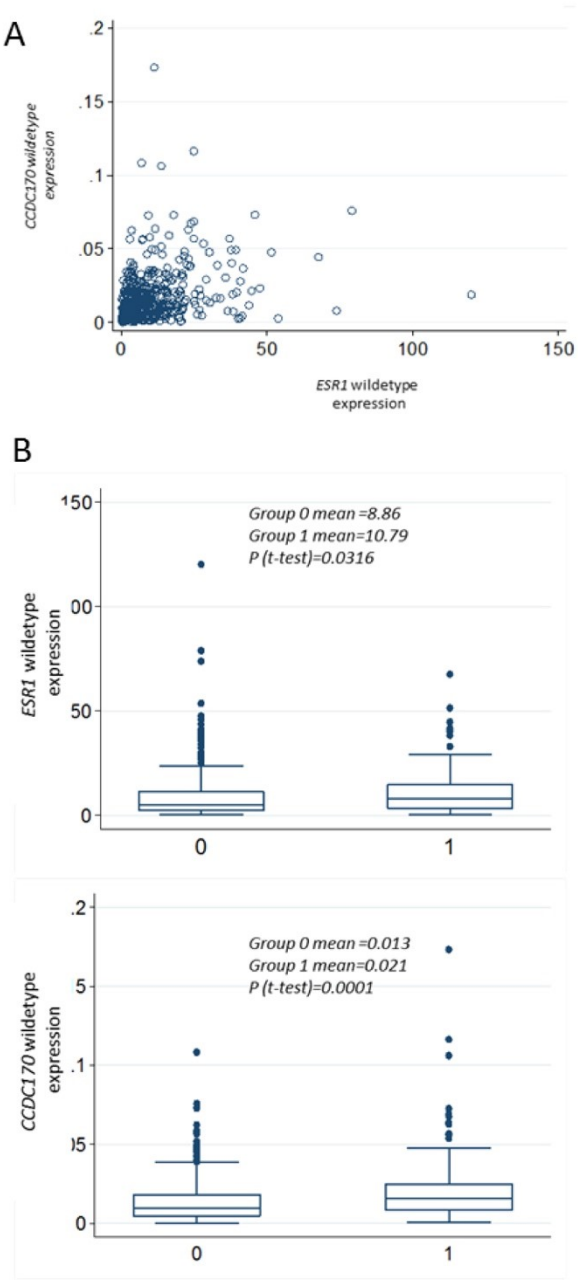


Figure S4. *ESR1* and *CCDC170* wildtype expression in ER-positive tumors compared between *CCDC170* fusion-negative and -positive cases.

A. Correlation between *CCDC170* and *ESR1* wildtype expression measured by RT-qPCR. B. *CCDC170* and *ESR1* wildtype mRNA levels were measured by RT-qPCR in samples with *ESR1-CCDC170* fusion transcript and compared to the group without fusion transcript. The box plots show interquartile ranges (IQR) together with the median (black horizontal line) of the *ESR1* and *CCDC170* mRNA levels for the different conditions. Group 0: *CCDC170*-fusion negative cases (n=387); Group 1: *CCDC*170-fusion positive cases (n=159).

Chapter 8

General Discussion

General Discussion

Liquid biopsy represents a non-invasive tool for genomic profiling of tumors, which can help to depict the heterogeneity of primary and metastatic tumors and to define the optimal targeted treatment. Compared to tissue biopsies, liquid biopsies are less invasive, painful and risky to the patient since they only requires a draw of blood, or the collection of saliva, urine or other physiological fluids from a cancer patient. In the context of the personalized medicine, liquid biopsies can provide insight into the biology of the cancer *in situ*, helping clinicians in treatment choice, and assist in accurate monitoring for the development of acquired resistance. Ultimately, the use of liquid biopsies may improve clinical management, quality of life and long-term survival for patients with cancer.

In this thesis, I explored the potential benefits of using liquid biopsies and biomarkers determined therein to diagnosis and follow up cancer patients in terms of personalized medicine. I particularly focused on pre-analytical, analytical and post-analytical conditions and investigated the role of different liquid biopsy biomarkers, such as circulating cell free DNA (cfDNA) and extracellular vesicles (EVs). The next paragraphs will present the current state of the art of the use of liquid biopsies and my contribution to the field including the discussion of the challenges related to the use of such liquid biomarkers affecting standard operating procedures in the laboratory. I will also discuss the applications of liquid biopsies derived biomarkers into routine diagnostics for the management of cancer patients.

1. Circulating cell free DNA and tumor derived cell free DNA detection

In recent years, circulating cell free DNA (cfDNA) has gained attention for its use as a biomarker in numerous fields, such as prenatal testing [noninvasive prenatal testing (NIPT)] [1], diabetes, cardiovascular diseases, organ transplantation, autoimmune diseases, sepsis, trauma, and even sports medicine [2]. Also in the field of cancer, considerable progress has been being made [3-5]. However, to allow widespread use of the cfDNA analysis and circulating tumor DNA (ctDNA) in routine clinical care further improvements are needed. One of the main challenges is a lack of standard operating procedures regarding the use of these biomarkers. For this purpose, in 2017, experts in the field of cancer biomarker testing were invited to participate to the International Quality Network (IQN) for pathology, with the aim to summarize the current state of the art on

liquid biopsy testing and to reach consensus on standards procedures in cfDNA and ctDNA testing. The meeting highlighted the need for guidelines about cfDNA testing procedures and result's interpretation, as well as the necessity for external quality assessment programs [6]. It was further concluded that the quantity and quality of cfDNA and ctDNA might be affected by many factors resulting in pre-analytical, analytical and post-analytical variability. While errors can occur during all these steps, the pre-analytical phase raised the most concern with error rates ranging from 46% to 68.2% [3, 7]. Therefore, the potential effects of biological or technical factors affecting the assay results in the pre-analytical phase should be determined to guarantee a correct result from cfDNA-based molecular test [8, 9].

1.1 Pre-analytical conditions

Pre-analytical conditions include procedures of sample collection (e.g. type of blood collection tube), transport, processing and storage. All these pre-analytical conditions should be optimal and compatible with routine clinical testing with regard to their lead-time, and their cost-effectiveness.

One of the biggest challenges in cfDNA isolation and ctDNA detection therein is related, first, to the poor yield of these biomolecules. Second, ctDNA detection, within cfDNA isolates, is hampered by the predominant presence of contaminating genomic DNA derived from healthy cells, mostly white blood cells. Moreover, the median half-life of cfDNA in circulation ranges from 15 minutes to a few hours [10]. Over the years, both serum and plasma have been used as biological sources for cfDNA [3]. However, it has been shown that normal DNA derived from leukocyte lysis during clotting is higher in serum than in plasma [11]. This finding has been confirmed by the analysis of cfDNA content, cfDNA fragments length and variant allele frequencies of ctDNA derived from cancers patients [12-14]). Hence, plasma is now considered the main source of cfDNA.

The first and important pre-analytical step in sample preparation consists of the choice of the blood collection tube being used. EDTA, heparin, and citrate are the most commonly used anticoagulants. However, heparin interferes with quantitative PCR (qPCR) [15, 16] and is therefore advised against. Therefore, for a long time, EDTA tubes have been recommended as blood collection tubes to obtain plasma for cfDNA isolation and analysis. However, work from our lab [17] that I was also involved in, showed that EDTA tubes are only suitable in routine clinical

practice if they are processed within 4-6 hours as delays between blood withdrawal and plasma preparation results in higher background of healthy genomic DNA due to the lysis of blood leukocytes. Therefore, special cell-stabilizing blood collection tubes are now commercially available, such as Streck's Cell-Free DNA BCT tubes (Streck) and CellSave preservative tubes (Menarini), containing preservatives that prevents lysis of leucocytes for up to several days at room temperature. These tubes are then advised when blood sampling and processing is not performed in the same clinical center or when delay between venipuncture and plasma separation is expected to be more than 4 hours. In our mentioned study, we demonstrated that CellSave and BCT tubes are able to prevent release of DNA from leukocytes and thus prevent dilution of ctDNA keeping mutant allele frequencies stable up to 96 hours from blood withdrawal.

The second pre-analytical step consists of obtaining cfDNA from plasma as pure as possible without unnecessary genomic DNA contamination. This requires an additional procedure during plasma preparation. A double-centrifugation protocol, consisting of an initial slow centrifugation speed to separate plasma followed by a fast centrifugation step to clear cellular material, has been recommended to obtain plasma [18, 19]. In this context, in our previous study, we also used a 2-step plasma separation procedure with a first centrifugation at 1700g at 4 °C for 10 min followed by a second-one at 12000g at 4 °C for 10 min [17]. Plasma should be then stored in multiple aliquots to reduce future freeze–thaw cycles. To increase cfDNA purity and stability, we also advised to add dithiothreitol at a 5 mmol/L when plasma is defrosted from -80°C for the first time to prevent nucleic acid degradation. Similarly, plasma quality should be checked before every cfDNA isolation. One of the main parameters to be checked is the presence of traces of hemolysis, which gives plasma a visible orange/red color. A previous study showed compromised PCR amplification on cfDNA extracted from plasma displaying hemolysis [20]. For reasons above, in **Chapter 6,** we restricted our analysis only to plasma collected within 24 hours from blood draw, and excluded all samples in which hemolysis was evident.

The third pre-analytical step consists of the appropriate isolation method to extract cfDNA with the required analytical homogeneity, yield, rapidity, and practicality. The isolation method used should be compatible with one important limitation with regard to cfDNA research, which is its low concentration in plasma. Various DNA extraction methods are now available [8, 21, 22] and these can be distinguished based on the type of separation method being either anion exchange, silica membrane, or magnetic bead-based ctDNA isolation protocol. Today, the silica membrane-based QIAamp Circulating Nucleic Acid kit represents the "golden standard " to obtain the highest

median cfDNA yield from plasma derived from cancer patients. However, its manual and time-consuming protocol render this method unsuitable for high-throughput isolations. In **Chapter 3,** we therefore investigated the use of two automated systems for cfDNA extraction (Maxwell AS1480 and QIASymphony) with respect to yield and reproducibility as variations between methods might results in different cfDNA yield and quality. We demonstrated that the QIASymphony automated platform, in contrast to the Maxwell AS1480 system, performed comparable to the "gold standard'" QIAamp manual system. This might explained by the presence of proteinase K incubation step in the QS and QA protocols, which is absent is Maxwell method. The proteinase K could improve cfDNA yield by inhibiting nucleases and the release of protein-bound cfDNA. However, as the Maxwell system selectively enriched shorter cfDNA fragments than both the QiaSymphony and the QIAamp protocols, the lower cfDNA yield did not affect into a significant difference in detected variant allele frequency. Overall, compared to a manual system, the use of an automated platform in diagnostic routine reduces hands-on-time, increases simultaneous sample throughput while yielding similar cfDNA concentrations (for QIAsymphony only) without large variability quality, and with a similar ctDNA frequency (both Maxwell and QiaSymphony).

In a recent study, *Lampignano* et al. performed a multicenter comparison of the cfDNA extraction and quantification methods by using different pre-analytical and analytical workflows. The authors firstly generated mononucleosomal DNA (mnDNA) from lung cancer cell lines with known *TP53* mutations, which were then spiked into pools of plasma from healthy donors collected in Streck and PAXgene tubes. The mnDNA was used to evaluate the performance of 6 different cfDNA extraction methods, including the column-based QIAamp CN, the QIAamp MinElute ccfDNA kits and the automated beads-based QIAsymphony, Maxwell AX1115 and AS1480, and Chemagic protocols. Although the QIAamp CNA was confirmed as the "golden standard" method for the highest recovery of cfDNA levels, the automated QIAsymphony system provided concordant results for all the quantification methods used (Quibit, qPCR and dPCR), followed by the Maxwell AX1115 platform. Moreover, Bioanalyzer assessed the length distribution of the recovered fragments. Overall, an enrichment of small fragments (140 - 166 bp) was observed in eluates obtained from QIAsymphony and QIAamp MinElute kits. Moreover, a smaller quantity of intrinsic wild-type cfDNA fragments were isolated by the Maxwell AX1115 platform resulting in an in artificially increased VAF of *TP53* mutations [23]. As in our study the

MX system isolated lower yield of both mutant and wild-type molecules, this did not translated into a significant difference in detected VAF.

To obtain reliable results, cfDNA should be isolated from an enough quantity of plasma. However, this is not always possible, particularly when a retrospective study is performed in which only a limited amount of body fluid has been collected. For this reason, in **Chapter 6** we developed an optimal workflow to isolate and investigate cfDNA from a minute amount of plasma (200 µL), yielding a relatively lower amount of cfDNA (in our study, ranging between 2 and 200 ng). In our study, we managed to establish and use an unbiased uniplex and multiplex pre-amplification step, which enabled us to detect the four most commonly found hotspot mutations in the estrogen receptor gene (m*ESR1*). This was firstly confirmed by experiments with different starting volumes of plasma from the same healthy blood donors, with or without pre-amplification step. In our study, we noted few mutant copies in cfDNA from pre-amplified healthy blood donor samples, which were not identified in the matched not pre-amplified sample. These false positives were maybe introduced by the Taq polymerase activity. Therefore, to prevent them from being scored as positive, a stringent cutoff value at 1% was used for all of the four m*ESR1* studied, which was well above the average VAF + 2.58 x SD measured in 18 individual healthy blood donors.

1.2 Cell free DNA downstream analysis

Solid tumors are heterogeneous and can modify their molecular profile, particularly during disease progression or under treatment pressure. Analysis of a small piece of tissue biopsy may fail to provide the correct landscape of the tumor either at diagnosis or during progression. In this context, liquid biopsies help in profiling the whole tumor at diagnosis, as well as its evolution during disease progression. Plasma from cancer patients can be easily obtained, and downstream cfDNA analysis are usually user friendly. Moreover, ctDNA analysis can support therapy guidance and longitudinal disease monitoring [23, 24].

However, the cfDNA amount might be variable between patients or over time within a patient and the fraction derived from tumor (ctDNA) represents only a small proportion of all cfDNA. Highly sensitive and specific methods are therefore required to detect mutations, particularly when they are present at low variant allele frequency. In this context, digital PCR (dPCR) is a sensitive and quantitative method based on parallel amplification of up to millions of individual

DNA fragments enabling the detection of alterations from minute quantities of ctDNA (less than 0.1%) [25, 26]. Of note, the limit of this technology is that you have to know beforehand what mutations you are looking for and that only few aberrations can be assessed simultaneously. For this reason, techniques that are able to assess more mutations or can evaluate the whole genome are currently available by next-generation sequencing (NGS). However, NGS methods are often not as sensitive as dPCR methods and mutations can be missed, particularly when the total number of reads are low. The two technologies (dPCR and NGS) are therefore now combined to obtain data that are more reliable and digital PCR is usually used for the independent validation of NGS results as we demonstrated in **Chapter 3**. In this study, we investigated the *TP53* gene status in tissues and archived sera from 20 HGSOC patients using two NGS workflows (Ampliseq and Oncomine NGS platform) alone or combined with digital dPCR. Overall, we showed that Oncomine NGS and dPCR systems are able to measure mutations below 1% VAF and are more sensitive than the conventional Ampliseq NGS. In fact, mutations were found in 67% of patients by dPCR and Oncomine NGS compared to only 50% of patients by Ampliseq NGS. Similarly, in **Chapter 6** E*SR1* mutations obtained by using dPCR were compared with analysis by Oncomine-targeted NGS. Overall, a high concordance of results (80%) was observed between the two methods, with the exception of one *ESR1* D538G mutation (VAF: 1.9%) which was identified only by the dPCR.

1.3 Challenges related to the use of cell free DNA in cancer

Although promising, challenges remain when using ctDNA to characterize the mutation status of a tumor. These challenges are related, first, to define whether ctDNA mirrors the characteristic of a tumor, and whether mutations detected in plasma are truly tumor-derived. Second, challenges are related to distinguish in ctDNA the *driver mutations* that provide a selective growth advantage, and thus promote cancer development, from those that do not (*passenger mutations*).

1.3.1 Discrimination between tumor-related mutations and clonal hematopoiesis-derived mutations

While several groups have used ctDNA as biomarker to investigate tumor mutations, the downside of this technology is that results need to be carefully taken due to the somatic mosaicism of the tumor. Cells into the bloodstream can acquire genetic variants, which may not represent true tumor genotype due to the clonal hematopoiesis (CH) of indeterminate potential process (CHIP) [27, 28].

Overall, CH is an age-dependent mechanism that can also occur in elderly healthy individuals, generally older than 70 years [29-31]. This process might generate genetic variants in hematopoietic progenitor cells and lead to the formation of distinct subpopulations [27, 29, 32]. Approximately 30% of patients with solid cancer harbor CH mutations in their blood and the presence of these mutations, as well as their VAFs variations over time, might affect disease risk [31, 33-35]. Up to ~10% [36] of all CH mutations affects *TP53* gene, which is also one of the main driver oncogenes in solid tumors [27, 37].

Plasma cfDNA is a complex mixture of mutations derived from germline DNA and malignancy [38]. When tumor DNA-only sequencing is assessed, any variants identified can represent either somatic changes or germline variants retained in the tumor DNA [39]. Thus, analysis of paired germline and tumor DNA from the patient should be considered and recommended to avoid misinterpretation of results. The presence of CH mutations in blood, although often at very low VAFs [27], are therefore only detectable when using highly sensitive ctDNA detection and quantification systems. An interesting study by *Hu* et al. revealed that most *JAK2* mutations, some *TP53* mutations, and rare *KRAS* alterations detected in cfDNA from NSCLC patients are

hematopoietic-derived mutations [28]. For this reason, in **Chapter 3** we decided to firstly check tumor-specific *TP53* mutations in tissue by NGS and then investigated the ctDNA. Interestingly, 67% of patients harbored the missense *TP53* mutations in both primary tumor and serum highlighting that these aberrations are tumor-derived. Our study confirmed the role of point mutations in *TP53* gene measured in cfDNA as biomarker for ovarian cancer patients.

1.3.2 Identify tumor-driver mutations and therapy resistance mutations

One of the main advantage of using cfDNA/ctDNA is the fact that it might reflect the tumor mass at the time of blood draw, thus allowing it to be used as marker of response to a given therapy or to predict relapse in patients with multiple tumor types, including colon, breast and lung cancer [24]. For this to work, it is mandatory to know the type of mutations/aberrations present in the cancer of interest and the disease stage (e.g. primary tumor *vs* metastatic disease; therapy sensitive or resistant tumor) because not always such mutations are shared. A perfect example is represented by *ESR1* mutations in breast cancer. These aberrations are hardly found (3%) in primary breast tumor tissue and in baseline plasma samples [40, 41], but do occur with high prevalence in patients with advanced disease (from 40% to 55%) having received prior treatment with aromatase inhibitors [40-42]. As a result, *ESR1* mutations are readily detected in cfDNA from ER-positive patients with recurrent disease after one or multiple lines of endocrine treatment, as previously demonstrated [43] and observed by myself in this thesis (**Chapter 6**).

Together with *ESR1, PIK3CA* and *TP53* gene mutations are promising biomarkers for the management of breast cancer patients. According to the Cancer Genome Atlas (TCGA), *PIK3CA* and *TP53* are the two most frequently mutated oncogenes in primary BC with frequencies of 30-40% and <30%, respectively [44]. An interesting recent study by *Rodriguez et al*. sequencing 29 primary breast tumor biopsies (of which 86.2% ER-positive) revealed a mutation rate of 65.5% and 20.7% in *PIK3CA* and *TP53* genes, respectively. To corroborate with these results, we assessed in **Chapter 6** the feasibility of using ctDNA NGS (Oncomine) to identify actionable mutations and serial ctDNA monitoring in blood, also allowing us to distinguish between driver mutations and/or therapy resistance alterations. In line with *Rodriguez et al*., we found that mutations in *PIK3CA* and *TP53* (either in 28.5% patients) genes were most prominently present close to m*ESR1*. With regard to the study of tumor evolution during treatment, in one patient the *PIK3CA* H1047R driver mutation was detected with high VAF both at baseline and progression disease (15.28% *vs* 4.2%, respectively) whereas, likely therapy resistance related mutations

PIK3CA E726K and *ESR1* Y537N had low VAF at baseline but increase considerably at time of progression, respectively (*PIK3CA* E726K: 0.34% vs 3.9% and *ESR1* Y537N: 0% vs 4.07%). In this context, serial monitoring of ctDNA allele fraction might acts a surrogate of response to therapy to allow personalized treatment decisions.

2. Clinical application of cell free DNA in breast cancer

Breast cancer is the most commonly diagnosed cancer and the main cause of cancer-related death for women worldwide. Detection of signs of relapse at diagnosis and during follow-up is one of the main aims in breast cancer management. *Coombes* and collegues [45] showed that patient-specific ctDNA analysis, every 6 months for up to 4 years, was able to predict relapse with a lead-time of up to 2 years [45]. Similarly, monitoring of ctDNA during endocrine therapy has shown that amongst others mutations in the *ESR1* gene are being selected during endocrine therapy. Overall, the screening of aberrations in *ESR1*, whether point mutations and gene fusions, might be used to stratify patient in subsets with differential response to treatment. For example, in the BOLERO-2 study the presence of *ESR1* Y537S and D538G mutations was associated to decrease overall survival [46] while in the PALOMA-3 trial, treatment with palbociclib with fulvestrant improved PFS compared to fulvestrant alone for patients with *ESR1* mutant cancers [42]. Another interesting study by *Kruger* et al. investigated whether ctDNA characteristics might be correlated with differential progression-free survival (PFS) in MBC patients receiving everolimus plus exemestane. Overall, the authors showed that patients with low/no ctDNA load or < 3 hotspot mutations had a longer PFS and OS compared to patients with higher ctDNA load [47]. Similarly, a study by *Lee* et al. revealed that high ctDNA load in plasma of breast cancer patients predicts disease recurrence and a worse outcome [48].

Likewise, in **Chapter 6** we investigated whether a correlation between the presence of mutation in *ESR1* and treatment existed within two retrospectively patient cohorts that received chemotherapy or endocrine therapy for metastatic disease. Our data confirmed an enrichment of *mESR1* in the metastatic setting after having been treated with endocrine therapy. Interestingly, we also detected *mESR1* in a patient who received chemotherapy but whom had never received endocrine therapy. This finding might be explained by intrinsic tumor genetic instability that allows cancer cells to develop mutations under treatment pressure or by the presence of the mutations already in the primary tumor, as previously reported by *Wang* et al [49]. Alternative might be that chemo-induced ovarian function suppression, mimicking the AI-

induced estrogen deprivation status that may promote the onset of *ESR1* mutations in premenopausal patients.

Similar to *mESR1*, *ESR1* fusions are aberrations often detectable in primary breast cancer, particularly in patients with high-grade disease or with resistance to endocrine therapy [50]. Expression of *ESR1* gene-fusions in ER+ breast cancer cells increase cell proliferation and decrease sensitivity to tamoxifen [51]. Particularly, *ESR1* fusion genes containing the first six exons of *ESR1* fused to C-terminal of gene partners have been mostly observed in metastatic ER+ breast cancer resistant to endocrine therapy [52]. Among the different fusions, *ESR1-CCDC170* fusion transcripts are the most predominant in breast cancer. On the contrary to the fact that these variants are related to endocrine resistance, in our study (**Chapter 7)** we did not find correlation between the presence of any of these fusion transcripts and outcome to endocrine therapy, whether it concerned first line tamoxifen or an aromatase inhibitor. However, we did observe that presence of *ESR1-CCDC170* exon 8 gene fusion transcript predict shorter disease free survival and overall survival. Moreover, we found gene fusions expression also in normal appearing margins of cancerous specimens. This argues that a percentage of these fusions may be transcript read-throughs instead of fusion transcripts arising from gene fusions.

3. Extracellular vesicles application and challenge in liquid biopsy

Extracellular vesicles (EVs), such as exosomes and microvesicles, are released by different cell types in body fluids such as blood, urine, cerebrospinal fluid, breast milk and saliva [53, 54]. It is well known that EVs mediate intercellular communication between the cells they originated to cells of target tissues and play a crucial role in various physiological and pathological processes (e.g. neurodegenerative disease [55] and cancer [56]). Nowadays, it is well known that EVs are involved in tumorigenesis, proliferation, drug resistance, angiogenesis and the development of pre-metastatic niches [57-61]. This ability is mainly related to their content of nucleic acid, DNA and RNA [62]. Particularly, as carriers of functional RNA species resulting elsewhere in oncogenic proteins, EVs have attracted much attention as potential diagnostic, prognostic, and predictive biomarkers and therapeutic agents. However, the current lack of standardized procedures for their isolation, quantification, and analysis challenges their use. Therefore, the International Society for Extracellular Vesicles (ISEV) proposed the Minimal Information for Studies of Extracellular Vesicles ("MISEV") guidelines in 2014, which were updated in 2018 [53, 63]. The main aim of these recommendations was to sensitize researchers, as well as journal editors and

reviewers, to the minimal information required for EVs analyses, including pre-analytical procedures involving separation, isolation and enrichment methods, characterization techniques, as well as proper guidelines for studying their functionality.

EVs can be easily obtained from cell culture conditioned media, biological fluids and tissues. When culture media are used as source of EVs, all the culture conditions should be annotated, including density/confluency at harvest, culture volume, surface coatings, oxygen or other gas tensions, stimulation and other treatments used such as antibiotics and/or growth factors that can affect EV production and/or composition [63].

If EVs are instead obtained from body fluids (e.g. plasma), several pre-analytical parameters should be taken into account. Similar to cfDNA analyses, pre-analytical conditions include the choice of the blood collection tube and type of anticoagulant [64-66], the time to processing blood, mixing or agitation, temperature (storage and processing), the use of centrifugation or filtration procedures and degree of hemolysis [63].

To date, differential ultracentrifugation is the main technique used for EVs separation and concentration. However, several additional techniques or combinations of techniques have been currently being developed, such as flow filtration and variations thereon, field-flow fractionation (FFF), asymmetric flow field-flow fractionation (AFFF, A4F, or AF4), field-free viscoelastic flow, alternating current electrophoresis, variations on size exclusion chromatography (SEC), ion exchange chromatography, and microfiltration [63].

When isolated, EVs should be then quantified and characterized. Quantification can be performed by light scattering technologies, such as nanoparticle tracking analysis (NTA); by standard flow cytometry for larger EVs or high-resolution flow cytometry for smaller EVs [63]. Particles characterization can be done by using methods as membrane-based affinity binding step of the Exo(RNA)Easy kit, NTA, transmission electron microscopy (TEM) and immune assays [67, 68]. Characterization of EVs may be done by their protein composition. Surface markers include A33 antigen (GPA33), EPCAM [69], tetraspanins CD63, CD9, and CD81 [70]. The use of EVs in clinical practice is challenged by the fact that it is not easily possible to establish the exact percentage of EVs that are tumor-derived [71, 72]. To circumvent this issue, one possible approach to quantify EVs consists in using tumor specific somatic variations which can be detected at DNA or RNA level, or gene fusions [56]. This finding has already demonstrated for different cancer types, such as glioma, ovarian and prostate cancer [73]. An alternative is to use

specific tumor or cell type markers, such as EpCAM, an epithelial marker enriched in carcinoma cells.

In this context, as described in **Chapter 5,** we firstly set-up a workflow to study EVs from cell lines and then applied this to investigate EVs derived from healthy blood donors and patients with different types of cancer. Our study demonstrated that EDTA tubes are more suited for EVs isolated from blood processed in plasma up to 24 h after venipuncture than BCT and CellSave tubes. EVs derived from patients and HBDs were quantified by NTA, TEM and immuno-assays as well as by using cytosolic protein FLOT-1 and EpCAM levels. Moreover, we used the CD9 TRIFic assay to measure the CD9 transmembrane protein, which is known to be enriched up to 10 times in EVs compared to other particles. Higher CD9 levels were obtained from plasma of cancer patients compared to healthy donors. Moreover, we observed a time-dependent increase in CD9 levels, which suggests that EVs are released by (blood) cells during storage after blood drawn but before blood was processed into plasma. We suggested using a cancer/cell type specific protein (e.g. EpCAM) as target for enrichment of tumor derived EVs. Next, using gene expression profiling of cellular mRNA and EV-RNA, we demonstrated that EVs carry multiple gene transcripts, which cluster with their matched cellular counterpart suggesting that EV-RNA contain molecular (sub) type specific genes. A similar comparison between EVs from patients and healthy donors did not yield any differences; therefore, we speculated that the majority of EVs in plasma is derived from healthy tissue. By using a dPCR approach, we showed a lower sensitivity for mutation detection in plasma EV-RNA compared to cfDNA (15% *vs* 75%). This difference can be partially explained by differences for plasma used (1 mL *vs* 3 mL), differences in time of plasma storage and analyses, presence of mutant target transcripts in EVs or purity of cancer derived EVs. Despite increasing scientific and clinical interest, lack of a standardized method for tumor-EVs enrichment hinders reliable and comparable quantification of blood-derived EVs. In this context, we suggested to implement methods for enrichment of tumor-specifc EVs to enable the study of tumor-specific information in liquid biopsy over time.

4. Recommendations

This thesis provide insight into technical aspect and possible clinical relevance of the liquid biopsies biomarkers (cfDNA and EVs) in cancer patients. This knowledge has been combined to provide a roadmap how the proper use of liquid biopsies allows for biomarker profiling of tumors, patient's stratification and choice of suitable targeted therapy.

In my view, the use of liquid biopsies is a powerful method for (early) cancer detection and represent great promise for precision medicine. The main advantages of liquid biopsies are that they can be minimally invasively retrieved and can be measured repeatedly over time enabling analysis of tumor dynamics (evolution of clonal heterogeneity) during disease progression, e.g. while on a specific targeted treatment, and as such overcomes limitations of current tissue biopsies. With regard to cfDNA characterization, we should keep in mind that all the analytical phases (pre-analytical, analytical, and post-analytical) of the analysis are important and prone to errors. In this sense, first, blood collected in EDTA tubes should be processed in plasma within 24 hours from venipuncture. If this cannot be guaranteed (e.g. in multicenter clinical studies), CellSave or BCT tubes can be used instead for processing blood into plasma up to 96 hours. Second, the volume of blood plasma for cfDNA isolation should be >1 mL, as plasma input affects the mutation detection rate at a low VAF and the possibility to perform multiple analyses on the same sample. If this is not feasible (e.g. in retrospective collected material) an unbiased uni/multiplex pre-amplification step can be used to accurately quantify variants in cfDNA. Third, comparable cfDNA quality and quantity can be obtained by the column-based manual protocol (QiAmp) and the automated magnetic-bead-based methods (QiaSymphony and Maxwell). Therefore, these automated platforms can replace the more laborious manual method, especially when high-throughput cfDNA isolation is needed. Fourth, when dealing with cfDNA, we should keep in the mind that its levels could vary widely in cancer patients, as confirmed in our cohort of HGSOC patients, where they were much higher at diagnosis, than after chemotherapy and progression. Moreover, correlations between tumor burden and cfDNA yields are not always linear. Fifth, downstream analysis on cfDNA can be promptly and easily performed by dPCR and NGS. Among them, dPCR has the advantage that it requires far less starting material (down to 0.1 ng cfDNA versus at least 5 ng cfDNA for the targeted NGS in our *mESR1* workflow), is cheaper because it needs less consumables and allows monitoring of a specific mutation over time in follow-up studies (as showed for *mESR1* and *TP53* mutations).

Lastly, cfDNA allows for detection of *ESR1* mutations and real-time monitoring of therapy response in recurrent breast cancer. In the future, the knowledge of the m*ESR1* status may help

clinicians to offer m*ESR1*-positive patients different lines of endocrine therapy to further lengthen the progression-free survival time. However, it should be noted that detection of tumor-specific mutations in cfDNA is not always possible, as shown in our study where missense *TP53* point mutations present in the primary tumor were detected only in 67% cfDNA of patients.

The analyses of EVs can provide information on gene and protein expression, which cfDNA analyses cannot provide. For EV-RNA characterization, we should keep in mind that blood cannot be collected in CellSave nor BCT tubes and is restricted to EDTA tubes with a restricted processing time (24 hours). Tubes allowing stabilizing cell-free RNA (cfRNA), such as BCT-RNA tubes, are only recently available and might be better alternatives. To discriminate tumor-EVs from those released by the (normal) cellular sources we should use a method able to isolate, enrich and characterize cancer-specific EVs. This aim can be achieved by selecting specific-antigenic markers on the EVs surfaces (e.g. CD9 or EpCAM proteins). Unfortunately, the results obtained in this thesis regarding EV-RNA characterization on gene expression profile and mutation detection levels did not point towards direct clinical utility of EVs. First because we did not find a real difference between gene profiling of EVs from patients and those derived from HBDs. Second, because tumor-specific mutations in blood, were less often observed in EV-RNA than in matched cfDNA. Certainly, this does not mean that EVs characterization is futile for cancer patients, because EVs remains a substrate, similar to CTCs, in which characterization can be performed at multiple dimensions (RNA, DNA and protein). However, further studies are needed to determine whether mutant transcripts are detectable in EVs.

In conclusion, liquid biopsies analyses are likely to play a critical role in cancer diagnostics. However, future studies should focus standardizing the entire process; from sampling to data analyses to be sure that, results obtained in the biofluids are reproducible. In the future, different liquid biopsy derived biomarkers should be simultaneously analyzed to give the best prospect for patient's stratification and personalized treatment.

References

1. Pritchard, K.I., *Endocrine therapy: is the first generation of targeted drugs the last?* J Intern Med, 2013. **274**(2): p. 144-52.
2. Riggins, R.B., et al., *Pathways to tamoxifen resistance.* Cancer Lett, 2007. **256**(1): p. 1-24.
3. Osborne, C.K. and R. Schiff, *Mechanisms of endocrine resistance in breast cancer.* Annu Rev Med, 2011. **62**: p. 233-47.
4. Schiavon, G., et al., *Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer.* Sci Transl Med, 2015. **7**(313): p. 313ra182.
5. Pejerrey, S.M., et al., *The Impact of ESR1 Mutations on the Treatment of Metastatic Breast Cancer.* Horm Cancer, 2018. **9**(4): p. 215-228.
6. Beije, N., et al., *Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients.* Mol Oncol, 2018. **12**(1): p. 48-57.
7. Vitale, S.R., et al., *An Optimized Workflow to Evaluate Estrogen Receptor Gene Mutations in Small Amounts of Cell-Free DNA.* J Mol Diagn, 2019. **21**(1): p. 123-137.
8. Robinson, D.R., et al., *Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer.* Nat Med, 2011. **17**(12): p. 1646-51.
9. Veeraraghavan, J., et al., *Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers.* Nat Commun, 2014. **5**: p. 4577.
10. Giltnane, J.M., et al., *Genomic profiling of ER(+) breast cancers after short-term estrogen suppression reveals alterations associated with endocrine resistance.* Sci Transl Med, 2017. **9**(402).
11. Li, S., et al., *Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts.* Cell Rep, 2013. **4**(6): p. 1116-30.
12. Yoshihara, K., et al., *The landscape and therapeutic relevance of cancer-associated transcript fusions.* Oncogene, 2015. **34**(37): p. 4845-54.
13. Veeraraghavan, J., et al., *Recurrent and pathological gene fusions in breast cancer: current advances in genomic discovery and clinical implications.* Breast Cancer Res Treat, 2016. **158**(2): p. 219-32.
14. Sun, Y., et al., *Evaluation of potential regulatory function of breast cancer risk locus at 6q25.1.* Carcinogenesis, 2016. **37**(2): p. 163-168.

15. Li, L., et al., *Therapeutic role of recurrent ESR1-CCDC170 gene fusions in breast cancer endocrine resistance.* Breast Cancer Res, 2020. **22**(1): p. 84.
16. Park, C., et al., *Integrative molecular profiling identifies a novel cluster of estrogen receptor-positive breast cancer in very young women.* Cancer Sci, 2019. **110**(5): p. 1760-1770.
17. Lei, J.T., et al., *ESR1 alterations and metastasis in estrogen receptor positive breast cancer.* J Cancer Metastasis Treat, 2019. **5**.
18. Kanke, Y., et al., *Gene aberration profile of tumors of adolescent and young adult females.* Oncotarget, 2018. **9**(5): p. 6228-6237.
19. Simons, Y., et al., *Molecular characterization of luminal breast tumors in African American women.* Journal of Clinical Oncology, 2021. **39**(15_suppl): p. 550-550.
20. Sieuwerts, A.M., et al., *How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study.* Clin Cancer Res, 2005. **11**(20): p. 7311-21.
21. Sieuwerts, A.M., et al., *Concentrations of TIMP1 mRNA splice variants and TIMP-1 protein are differentially associated with prognosis in primary breast cancer.* Clin Chem, 2007. **53**(7): p. 1280-8.
22. Ramirez-Ardila, D.E., et al., *Hotspot mutations in PIK3CA associate with first-line treatment outcome for aromatase inhibitors but not for tamoxifen.* Breast Cancer Res Treat, 2013. **139**(1): p. 39-49.
23. Hayward, J.L., et al., *Assessment of response to therapy in advanced breast cancer. A project of the programme on clinical oncology of the International Union against Cancer, Geneva, Switzerland.* Eur J Cancer, 1978. **14**(11): p. 1291-2.
24. Onstenk, W., et al., *Efficacy of Cabazitaxel in Castration-resistant Prostate Cancer Is Independent of the Presence of AR-V7 in Circulating Tumor Cells.* Eur Urol, 2015. **68**(6): p. 939-45.
25. Sieuwerts, A.M., et al., *Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR.* Breast Cancer Res Treat, 2009. **118**(3): p. 455-68.
26. Sieuwerts, A.M., et al., *mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients.* Clin Cancer Res, 2011. **17**(11): p. 3600-18.

27. Lei, J.T., X. Gou, and M.J. Ellis, *ESR1 fusions drive endocrine therapy resistance and metastasis in breast cancer.* Mol Cell Oncol, 2018. **5**(6): p. e1526005.
28. Paratala, B.S., et al., *Emerging Role of Genomic Rearrangements in Breast Cancer: Applying Knowledge from Other Cancers.* Biomark Cancer, 2016. **8**(Supple 1): p. 1-14.
29. Jones, D.T., et al., *Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas.* Cancer Res, 2008. **68**(21): p. 8673-7.
30. Lipson, D., et al., *Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies.* Nat Med, 2012. **18**(3): p. 382-4.
31. Kim, R.N., et al., *Perspective Insight into Future Potential Fusion Gene Transcript Biomarker Candidates in Breast Cancer.* Int J Mol Sci, 2018. **19**(2).
32. Viale, G., et al., *Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1-98.* J Clin Oncol, 2007. **25**(25): p. 3846-52.
33. Hede, K., *Breast cancer testing scandal shines spotlight on black box of clinical laboratory testing.* J Natl Cancer Inst, 2008. **100**(12): p. 836-7, 844.
34. Fakhri, G.B., et al., *Concordance between Immunohistochemistry and Microarray Gene Expression Profiling for Estrogen Receptor, Progesterone Receptor, and HER2 Receptor Statuses in Breast Cancer Patients in Lebanon.* Int J Breast Cancer, 2018. **2018**: p. 8530318.
35. Roodi, N., et al., *Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer.* J Natl Cancer Inst, 1995. **87**(6): p. 446-51.
36. Brinkman, J.A. and D. El-Ashry, *ER re-expression and re-sensitization to endocrine therapies in ER-negative breast cancers.* J Mammary Gland Biol Neoplasia, 2009. **14**(1): p. 67-78.
37. Babiceanu, M., et al., *Recurrent chimeric fusion RNAs in non-cancer tissues and cells.* Nucleic Acids Res, 2016. **44**(6): p. 2859-72.
38. Martelli, M.P., et al., *EML4-ALK rearrangement in non-small cell lung cancer and non-tumor lung tissues.* Am J Pathol, 2009. **174**(2): p. 661-70.
39. Maes, B., et al., *The NPM-ALK and the ATIC-ALK fusion genes can be detected in non-neoplastic cells.* Am J Pathol, 2001. **158**(6): p. 2185-93.
40. Li, H., et al., *A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells.* Science, 2008. **321**(5894): p. 1357-61.

41. Biernaux, C., et al., *Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals.* Blood, 1995. **86**(8): p. 3118-22.
42. Li Li, J.V., Yiheng Hu, Xian Wang, Ying Tan, Rachel Schiff and Xiaosong Wang, *Abstract 376: Therapeutic role of ESR1-CCDC170 gene fusion in breast cancer endocrine resistance.* American Association for Cancer Research, 2019. **79**(13): p. Supplement, pp. 376.
43. Varley, K.E., et al., *Recurrent read-through fusion transcripts in breast cancer.* Breast Cancer Res Treat, 2014. **146**(2): p. 287-97.

Chapter 9

English Summary/ Samenvatting/ Sommario

English Summary

The studies that I have described in this thesis focus on improvement of the prognostic and predictive value of circulating biomarkers (ctDNA and Extracellular vesicles) by the optimization of the pre-analytical and post-analytical conditions as well as the choice of the suitable method to investigate rare mutations in biofluids of patients with different types of tumors. In **Chapter 3** we compared the performance of two automated cfDNA isolation systems, namely Maxwell (MX) and QIAsymphony (QS), to the the QIAamp Circulating Nucleid Acid (QA) Kit manual extraction kit, which is one of the most widespread methods (i.e. the golden standard) for cfDNA isolation. To date, several studies have compared different cfDNA extraction methods and have indeed demonstrated that the isolation method can considerably affect cfDNA yield. We thus decided to assess several cfDNA parameters (e.g. yield, recovery efficiency, fragmentation patterns, and ctDNA fraction retrieved) using optimally processed plasma samples of healthy blood donors (HBDs) and patients with metastatic cancer to establish the suitable isolation method for large clinical trial. We firstly investigated the effect of carrier RNA (cRNA) addition to cfDNA yield obtained from HBD plasma in the automated bead-based platforms. Although Qubit measurement revealed an increased cfDNA yield, this result was not confirmed by data from plant DNA qPCR, TERT and fragmentation dPCR assays. In a prior project, our group has shown that CellSave tubes ensure optimal ctDNA quality in plasma processed within 96 h after venipuncture compared to EDTA tubes which need to be processed with 24 h. Therefore, we tested the compatibility of these collection tubes with QS and MX and found that they do not affect the cfDNA downstream analysis. We then investigated the recovery efficiency in terms of fragment size of plant DNA spiked-in and showed that MX performed worse than QA and QS. Moreover, as cfDNA isolation procedure can affect the recovery of shorter and longer cfDNA fragments, it is plausible that tumor mutation quantification could be affected by the extraction method. We therefore investigated the somatic variant detection in ctDNA isolated by using different platforms. QS results were most comparable to QA while MX detected fewer mutant and wild-type molecule. However, this difference did not result in a significantly different VAF between all 3 methods compared.

Higher levels of cfDNA isolated from blood of patients with ovarian cancer patients have been associated to an advanced disease stage, high grade, and worse prognosis. Moreover, 96% of high-grade serous ovarian carcinomas (HGSOC) harbor mutations on *TP53* gene. In **Chapter 4** we therefore investigated the diagnostic, predictive and prognostic value of *TP53* missense

mutations measured in tissue and in archived serum cfDNA of 20 HGSOC patients analyzed by targeted NGS alone or in combination with dPCR. For 10 HGSOC patients, cfDNA was isolated from sera taken at different time points (diagnosis, during chemotherapy and at disease progression). The cfDNA yield obtained was widely variable, although it was mostly (55% of cases) above 100 ng/mL at diagnosis. On the contrary, cfDNA yield from blood taken during chemotherapy and at progression was lower than 100 ng/mL in almost all cases. Several studies have shown that cfDNA mutational status may reflect the genetic characteristics of the primary or metastatic lesion from which it is derived. This finding was confirmed by our study as we were able to detect the missense *TP53* point mutation present in the primary tumor in cfDNA from 67% of patients. Overall unimolecular barcoded NGS and dPCR assays were more sensitive than the conventional Amplicon-baesd NGS, enabling detection of *TP53* mutations below 1% allele frequencies. Moreover, dPCR allowed the longitudinal monitoring of these variations through treatment. We additionally confirmed the role of cfDNA as early molecular response marker for HGSOC patients by identifying *TP53* mutations at disease progression, but not during chemotherapy treatment.

Even though cfDNA released by tumor cells is considered to preserve the characteristics of the tissue of origin, cfDNA analyses does not provide information on gene and protein expression. In this context, Extracellular vesicles (EVs) have emerged as important liquid biopsy biomarker. The main hypothesis regarding EVs is that they are released into body fluids during various pathological processes, including cancer and that they contribute to intercellular communications even over longer distances. In the study described in **Chapter 5** we set up a workflow to isolate and characterize EVs according to MISEV2018-guidelines. This method was firstly tested in cell line models and then applied to analyze EV-RNA from plasma of 20 patients with metastatic cancer. The 96 gene expression profiles by RT-qPCR revealed a high concordance between cellular mRNA and EV-RNA as well as in VAF measured by dPCR for *PIK3CA*, *KRAS* and *BRAF* genes suggesting that EVs mirror their matched cellular counterpart. As several studies have shown that EV downstream analysis may be affected by the choice of type of blood collection tube, we decided to investigate quality and quantity of EVs from EDTA, BCT, and CellSave tubes. Although EDTA tubes seems to be the mostly suitable tube, for EVs isolation and EV-RNA extraction the impact of time delay between blood sampling and plasma processing must be taken into account. In fact, we observed increased number of copies of *GAPDH* and beta-2-microglobulin (*B2M*) derived from white blood cells with time to processing blood into plasma

after blood draw. However, since the increase within one day was not significant, we recommended for all future clinical studies to isolate EVs from blood collected in EDTA tubes within 24 hours. Previous studies have detected tumor-specific mutations in EV-RNA. We thus decided to investigate by dPCR EV-RNA mutational status of target genes with a known somatic variant in tissue. Our procedure was able to detect mutations in only 3/20 (15%) patients. This low detection rate might be due to the plasma amounts used as input for EV-RNA isolation (less than 1 mL) and/or to the necessity of enrichment for cancer- and/or tissue-specific EVs before genotyping oncogenic transcripts therein.

About 14-39% of metastatic breast cancer (MBC) patients acquire resistance to endocrine therapy due to the presence of mutations in the gene coding for ER (*ESR1*). Most of these mutations (*mESR1*) cluster at a hotspot region within the ligand binding domain of *ESR1*, resulting in an amino acid substitution at position 537 and 538 in helix 12 of the ER. Consequently, reliable detection of *mESR1* in cfDNA from plasma of MBC patients play an important role in future clinical treatment decision making. Although easily accessible in biofluids, cfDNA is often present at very low levels. Therefore, in **Chapter 6** we set-up and optimize a reliable and reproducible method to study m*ESR1* in minute amounts (200 µL) of plasma and cfDNA (2 µL) from MBC patients. To amplify in a targeted manner the cfDNA isolated (yield: 2-200 ng), we set-up a target-specific (uniplex or multiplex-targets) pre-amplification step and used a stringent cut-off value at 1% for all the *ESR1* mutants to be detected to prevent false-positives. Results obtained by using dPCR were then compared to data from targeted NGS covering hotspot mutations in 10 genes including *ESR1*. Our analysis confirmed previous studies showing that m*ESR1* are enriched in MBC patients at progressive disease compared to before the start of endocrine therapy (34.1% vs 3.9%, respectively). Additionally, we showed that *mESR1* can arise in the absence of hormone deprivation as it is not restricted to patients having received endocrine treatment. In fact, we were able to detect *mESR1* also in three subjects treated with chemotherapy and who had never received endocrine therapy.

Gene fusions have been considered an oncogenic event in several haematological and solid tumors, including breast cancer. Here, genetic aberrations are preferentially detected in primary breast cancer, particularly in the more aggressive forms, such as luminal B, basal like, or endocrine resistant breast cancer. Moreover, their functional characterization highlighted their

potential role in tumorigenesis and therapeutic resistance. In breast cancer, recurrent gene fusions have been reported between *ESR1* and partners genes *CCDC170, AKAP12 and ARMT1*, resulting in chimeric proteins with a constitutive hyperactivity that might be associated with resistance to endocrine therapy. For this reason, in **Chapter 7** we decided to investigate the occurrence of such genomic rearrangements and correlated their presence with clinical outcome in breast cancer patients. Our analysis revealed that *ESR1* fusion genes, particularly *ESR1–CCDC170*, are frequent events in breast cancer cell lines and in ESR1+ breast cancer subjects but less so in ESR1- cases (frequency: 25.8% vs 2.3%, respectively). Recurrent genomic rearrangements were mainly identified between the *ESR1* and the *CCDC170* locus, but particularly involved exons 2, 6 and 8. Interestingly, expression of these rearrangements was also measured in fibroadenoma and ductal carcinomas in situ (DCIS) tissues from women without the diagnosis of breast cancer as well as in normal adjacent mammary epithelial tissues of breast cancer patients. Overall, expression of the *ESR1-CCDC170* exon 2-exon 8 fusions were associated to a shorter disease-free survival, worse prognosis and shorter overall survival but not with response to 1st line endocrine therapy.

In conclusion, in this thesis we highlighted the need for standardized and optimized procedures for clinical cfDNA and EVs testing and reporting. Moreover, we showed how the use of liquid biopsy as well as the discover of more convincing biomarkers may help to better understand the evolving genomic landscape of solid tumors and treat different clusters of cancer patients. Overall, this twofold approaches have the potential for broader applications in clinical practice in the future by improving our ability to diagnose, predict recurrence, metastasis and drug resistance in cancer patients.

Nederlandse Samenvatting

De studies die ik in dit proefschrift heb beschreven, richten zich op het verbeteren van de prognostische en voorspellende waarde van circulerende biomarkers (ctDNA en extracellulaire blaasjes) door de optimalisatie van de pre-analytische en post-analytische condities, evenals de keuze van de geschikte methode om zeldzame mutaties in lichaamsvloeistoffen van patiënten met verschillende soorten tumoren te onderzoeken. In **hoofdstuk 3** vergeleken we de prestaties van twee geautomatiseerde cfDNA-isolatiesystemen, namelijk Maxwell (MX) en QIAsymphony (QS), met de QIAamp Circulating Nucleid Acid (QA) Kit handmatige isolatiemethode, een van de meest gebruikte methode (d.w.z. de gouden standaard) voor cfDNA-isolatie. Tot op heden hebben verscheidene studies verschillende cfDNA-extractiemethoden vergeleken en hebben ze inderdaad aangetoond dat de isolatiemethode de cfDNA-opbrengst aanzienlijk kan beïnvloeden. We hebben daarom besloten om verschillende cfDNA-parameters (bijv. opbrengst, isolatie efficiëntie, fragmentatiepatronen en geisoleerde ctDNA-fractie) te beoordelen met behulp van optimaal verwerkte plasmamonsters van gezonde bloeddonoren (HBD's) en patiënten met gemetastaseerde kanker om de geschikte isolatiemethode voor klinische studies vast te stellen. We onderzochten eerst het effect van drager RNA (cRNA) toevoeging aan cfDNA-opbrengst verkregen uit HBD-plasma in de geautomatiseerde isolatie platforms. Hoewel Qubit-meting een verhoogde cfDNA-opbrengst aan het licht bracht, werd dit resultaat niet bevestigd door gegevens van planten-DNA qPCR-, TERT- en fragmentatie-dPCR-assays. In een eerder project heeft onze groep aangetoond dat CellSave-buizen voor een optimale ctDNA-kwaliteit zorgen in plasma dat binnen 96 uur na bloedafname wordt verwerkt in vergelijking met EDTA-buizen die binnen 24 uur moeten worden verwerkt. Daarom hebben we de compatibiliteit van deze opvangbuizen met QS en MX getest en ontdekten we dat ze geen invloed hebben op de cfDNA downstream-analyse. Vervolgens onderzochten we de isolatie efficiëntie in termen van fragmentgrootte van plant-DNA en toonden aan dat MX slechter presteerde dan QA en QS. Aangezien de cfDNA-isolatieprocedures de isolatie efficiëntie van kortere en langere cfDNA-fragmenten kan beïnvloeden, is het aannemelijk dat de kwantificering van tumormutaties door de extractiemethode kan worden beïnvloed. Daarom onderzochten we de somatische variantdetectie in ctDNA geïsoleerd met behulp van verschillende platforms. QS-resultaten waren het meest vergelijkbaar met QA, terwijl MX minder mutante en wildtype moleculen detecteerde. Dit verschil resulteerde echter niet in een significant andere Variant Allel Frequentie (VAF) tussen alle 3 de vergeleken methoden.

Hogere niveaus van cfDNA geïsoleerd uit bloed van patiënten met eierstokkanker patiënten zijn geassocieerd met een vergevorderd ziektestadium, hooggradige ziekte en slechtere prognose. Bovendien herbergt 96% van de HoogGradig Sereuze Ovariële Carcinomen (HGSOC) mutaties op het TP53-gen. In **hoofdstuk 4** onderzochten we daarom de diagnostische, voorspellende en prognostische waarde van TP53 puntmutaties gemeten in weefsel en in gearchiveerd serum cfDNA van 20 HGSOC patiënten geanalyseerd door panel sequencing (panel NGS) alleen of in combinatie met digitale PCR (dPCR). Bij 10 HGSOC-patiënten werd cfDNA geïsoleerd uit sera die op verschillende tijdstippen werden afgenomen (diagnose, tijdens chemotherapie en bij ziekteprogressie). De verkregen cfDNA-opbrengst was zeer variabel, hoewel deze bij de diagnose meestal (55% van de gevallen) boven de 100 ng/ml lag. Integendeel, cfDNA-opbrengst uit bloed genomen tijdens chemotherapie en bij progressie was in bijna alle gevallen lager dan 100 ng/ml. Verschillende studies hebben aangetoond dat cfDNA-mutatiestatus de genetische kenmerken kan weerspiegelen van de primaire of gemetastaseerde laesie waaruit deze is afgeleid. Deze bevinding werd bevestigd door onze studie omdat we de TP53 puntmutatie in de primaire tumor ook in cfDNA konden detecteren bij 67% van de patiënten. Over het algemeen waren op moleculair barcode gebaseerde NGS-assays en dPCR-assays gevoeliger dan de conventionele NGS, waardoor TP53-mutaties onder 1% allelfrequenties konden worden gedetecteerd. Bovendien maakte dPCR de longitudinale monitoring van deze variaties gedurende behandeling mogelijk. We bevestigden bovendien de rol van cfDNA als vroege moleculaire responsmarker voor HGSOC-patiënten door TP53-mutaties te identificeren bij ziekteprogressie, maar niet tijdens chemotherapiebehandeling.

Hoewel cfDNA dat door tumorcellen wordt vrijgegeven, wordt geacht de kenmerken van het weefsel van oorsprong te behouden, bieden cfDNA-analyses geen informatie over gen- en eiwitexpressie. In deze context zijn extracellulaire blaasjes (EV’s) naar voren gekomen als belangrijke vloeistof biopt biomarker. De belangrijkste hypothese met betrekking tot EV's is dat ze vrijkomen in lichaamsvloeistoffen tijdens verschillende pathologische processen, waaronder kanker en dat ze bijdragen aan intercellulaire communicatie, zelfs over langere afstanden. In de studie beschreven in **hoofdstuk 5** hebben we een workflow opgezet om EV's te isoleren en te karakteriseren volgens MISEV2018-richtlijnen. Deze methode werd eerst getest in cellijnmodellen en vervolgens toegepast om EV-RNA te analyseren uit plasma van 20 patiënten met gemetastaseerde kanker. De 96 genexpressieprofielen van RT-qPCR onthulden een hoge concordantie tussen cellulair mRNA en EV-RNA, evenals in VAF gemeten door dPCR voor PIK3CA, KRAS- en BRAF-genen, wat suggereert dat EV's hun overeenkomende cellulaire tegenhanger

weerspiegelen. Omdat verschillende studies hebben aangetoond dat EV downstream-analyse kan worden beïnvloed door de keuze van het type bloedafnamebuis, hebben we besloten om de kwaliteit en kwantiteit van EV's van EDTA-, BCT- en CellSave-buizen te onderzoeken. Hoewel EDTA-buizen de meest geschikte buis lijken te zijn, moet voor EV-isolatie en EV-RNA-extractie rekening worden gehouden met de impact van de tijdfactor tussen bloedafname en plasmaverwerking. We zagen een verhoogd aantal kopieën van *GAPDH* en bèta-2-microglobuline (*B2M*) in witte bloedcellen welke gerelateerd was met de tijd om bloed in plasma te verwerken na bloedafname. Aangezien de toename binnen één dag echter niet significant was, is er voor toekomstige klinische studies aanbevolen om EV's binnen 24 uur te isoleren van bloed dat in EDTA-buizen wordt verzameld. Eerdere studies hebben tumorspecifieke mutaties in EV-RNA gedetecteerd. We hebben daarom EV-RNA mutatiestatus met dPCR onderzocht van doelgenen met een bekende somatische variant in weefsel. Onze procedure was in staat om mutaties te detecteren bij slechts 3/20 (15%) patiënten. Deze lage detectieniveau kan te wijten zijn aan de plasmahoeveelheden die worden gebruikt als input voor EV-RNA-isolatie (minder dan 1 mL) en/of aan de noodzaak van verrijking voor kanker- en/of weefselspecifieke EV's voordat er oncogene transcripties in kunnen worden gegenotypeerd.

Ongeveer 14-39% van de gemetastaseerde borstkanker (MBC) patiënten krijgen resistentie tegen endocriene therapie als gevolg van de aanwezigheid van mutaties in het gen coderend voor de estrogeen receptor (*ESR1*). De meeste van deze mutaties (*mESR1*) clusteren in een hotspotgebied binnen het ligand bindende domein van *ESR1*, wat resulteert in een aminozuursubstitutie op positie 537 en 538 in helix 12 van de ER. De betrouwbare detectie van *mESR1* in cfDNA uit plasma van MBC-patiënten speelt een belangrijke rol in toekomstige beslissingen over de klinische (vervolg)behandeling. Hoewel gemakkelijk toegankelijk in lichaamsvloeistoffen, is cfDNA en ctDNA vaak aanwezig op zeer lage niveaus. Daarom hebben we in **hoofdstuk 6** een betrouwbare en reproduceerbare methode opgezet en geoptimaliseerd om m*ESR1* in minieme hoeveelheden (200 µL) plasma en cfDNA (2 µL) van MBC-patiënten te bestuderen. Om de geïsoleerde cfDNA (opbrengst: 2-200 ng) ter verrijken, hebben we een doelspecifieke (uniplex- of multiplex-targets) pre-amplificatiestap opgezet en een strikte cut-off waarde van 1% gebruikt voor alle *ESR1* mutanten die konden worden gedetecteerd om vals-positieven te voorkomen. De resultaten verkregen door het gebruik van dPCR werden vervolgens vergeleken met gegevens van gerichte NGS met betrekking tot hotspotmutaties in 10 genen, waaronder *ESR1*. Onze analyse bevestigde eerdere studies waaruit bleek dat m*ESR1* verrijkt is bij MBC-patiënten met progressieve ziekte in

vergelijking met vóór het begin van endocriene therapie (respectievelijk 34,1% versus 3,9%). Bovendien hebben we laten zien dat *mESR1* kan ontstaan in patiënten die geen endocriene behandeling hebben gekregen, namelijk *mESR1 werd ge*detecteerd bij drie personen die met chemotherapie werden behandeld en die nooit endocriene therapie hadden gekregen.

Genfusies worden beschouwd als een oncogene gebeurtenis in verschillende hematologische en solide tumoren, waaronder borstkanker. Hier worden genetische afwijkingen bij voorkeur gedetecteerd bij primaire borstkanker, met name in de agressievere vormen, zoals luminale B, basale of endocriene resistente borstkanker. De functionele karakterisering van genfusies toonde hun potentiële rol in tumorigenese en therapeutische resistentie aan. Bij borstkanker zijn terugkerende genfusies gemeld tussen *ESR1*- en partnersgenen *CCDC170, AKAP12 en ARMT1*, wat resulteert in chimerische eiwitten met een constitutieve hyperactiviteit die geassocieerd kan worden met resistentie tegen endocriene therapie. Om deze reden hebben we in **hoofdstuk 7** besloten om het optreden van dergelijke genomische herschikkingen te onderzoeken en hun aanwezigheid te correleren met klinische resultaten bij borstkankerpatiënten. Uit onze analyse bleek dat ESR1-fusiegenen, met name *ESR1–CCDC170*, frequent voorkomen in borstkankercellijnen en bij ESR1-positieve borstkankerpatiënten, maar minder in ESR1-negatieve gevallen (frequentie: respectievelijk 25,8% versus 2,3%). Terugkerende genomische herschikkingen werden voornamelijk vastgesteld tussen de *ESR1* en de *CCDC170* locus, maar in het bijzonder met exon 2, 6 en 8 van CCDC170. Interessant is dat de expressie van deze herschikkingen ook werd gemeten in fibroadenoom en ductale carcinomen in situ (DCIS) weefsels van vrouwen zonder de diagnose van borstkanker en in normale aangrenzende borstepitheelweefsels van borstkankerpatiënten. Over het algemeen werden expressies van de *ESR1-CCDC170* exon 2-exon 8 fusies geassocieerd met een kortere ziektevrije overleving, slechtere prognose en kortere algehele overleving, maar niet met 1e lijn endocriene therapie resistentie.

Tot slot, in dit proefschrift hebben we gewezen op het belang van gestandaardiseerde en geoptimaliseerde procedures voor het testen en rapporteren van klinische cfDNA en EV's. We hebben laten zien hoe het gebruik van vloeistof biopten en de ontdekking van meer overtuigende biomarkers kunnen helpen om het evoluerende genomische landschap van solide tumoren beter te begrijpen en verschillende clusters van kankerpatiënten te behandelen. Over het algemeen hebben deze tweeledige benaderingen het potentieel voor bredere toepassingen in de klinische

praktijk in de toekomst door ons vermogen om recidief, uitzaaiingen en resistentie tegen geneesmiddelen bij kankerpatiënten beter te diagnosticeren en te voorspellen.

Italian Sommario

La recente scoperta di alterazioni genetiche responsabili, in alcuni casi, della progressione tumorale, ha portato alla formulazione di terapie a bersaglio molecolare. Lo studio descritto in questa tesi ha l'obiettivo principale di definire ed ottimizzare le condizioni pre-analitiche ed analitiche, cosi come la scelta delle metodologie più idonee per migliorare il valore predittivo e prognostico di biomarcatori circolanti [DNA libero tumorale (ctDNA) e vescicole extracellulari (EVs)] utili per studiare alterazioni genetiche rare in biofluidi ottenuti da pazienti affetti da differenti forme tumorali.

A tal fine, nel **Capitolo 3** abbiamo confrontato la capacità di due sistemi di estrazione di acidi nucleici automatici [Maxwell (MX) e QIAsymphony (QS)] di isolare il DNA libero circolante (cfDNA), rispetto al sistema manuale ad oggi più utilizzato che si avvale dell'uso del kit "QIAamp Circulating Nucleid Acid" (QA). Per confrontare i tre sistemi di estrazione abbiamo valutato diversi parametri, fra cui la quantità e qualità di cfDNA estratto, la grandezza, in ordine di paia di basi (bp), dei frammenti di cfDNA isolato e la capacità di recuperare la frazione di ctDNA. Tale analisi è stata inizialmente condotta usando plasma isolato da sangue di donatori sani e in seguito su plasma di pazienti affetti da diverse forme di carcinoma.
Dapprima abbiamo valutato se l'aggiunta di un reagente, il "carrier RNA", migliora la resa di cfDNA isolato dalle due piattaforme automatiche. Nonostante la quantizzazione con spettrofluorimetro (Qubit) avesse rivelato un incremento della concentrazione del cfDNA estratto dai due sistemi automatici, questo dato non è stato confermato da successive analisi basate sull'utilizzo di metodiche più sensibili, quali la Real-Time PCR (RT-qPCR) e la Digital PCR (dPCR). Con la prima tecnica abbiamo ricercato la sequenza di un DNA sintetico derivato dalle piante inserito artificialmente nel plasma di donatori sani. La seconda metodica è stata invece usata per ricercare il gene controllo *TERT* e il tipo di frammenti (in bp) di cfDNA isolato.
Abbiamo poi voluto studiare se le provette CellSave sono compatibili con le piattaforme MX e QS, poiché in un nostro lavoro precedente avevamo dimostrato che il sangue raccolto in questi tubi può essere conservato fino ad un massimo di 96 ore (a differenza di quello racconto in tubi EDTA). La nostra analisi ha: *i)* confermato l'applicabilità dei tubi CellSave con le due piattaforme QS e MX; *ii)* dimostrato che le prestazioni del metodo QS sono simili a quelle di QA e migliori di

MX; *iii)* evidenziato che, nonostante le tre piattaforme differiscano per la grandezza dei frammenti di cfDNA isolati (in bp), questo dato non si traduce in una differente variante di frequenza allelica (VAF) delle mutazioni riscontrate dalle tre piattaforme di estrazione

Diversi studi hanno dimostrato che elevati livelli di cfDNA ottenuto da sangue di pazienti con carcinoma ovarico, sono stati associati ad una patologia ovarica in fase avanzata, di alto grado e con prognosi sfavorevole. Poiché circa il 96% dei pazienti con carcinoma sieroso ovarico di alto grado (HGSOC) ha mutazioni a carico del gene *TP53*, nel **Capitolo 4** abbiamo scelto di studiare il valore diagnostico, predittivo e prognostico delle alterazioni a carico di *TP53*. Lo studio è stato condotto usando il tessuto e il siero di 20 pazienti affetti da HGSOC attraverso il Sequenziamento di Nuova Generazione (NGS). I risultati ottenuti dal sequenziamento NGS sono stati inoltre comparati e/o validati mediante la tecnica di dPCR. A tal fine, il cfDNA è stato isolato da siero ottenuto da 10 donne con HGSOC in tre momenti diversi: alla diagnosi, durante la chemioterapia ed al momento della progressione della malattia. Benché la quantità di cfDNA ottenuto sia stata abbastanza variabile, la sua concentrazione è stata quasi sempre (nel 55% dei casi) superiore a 100 ng/mL per i campioni di pazienti alla diagnosi.
E' Inoltre noto che il ctDNA rispecchia le caratteristiche patologiche del tumore, primario o della lesione metastatica, da cui si origina. Di conseguenza, abbiamo voluto investigare se nel ctDNA possono essere identificate alterazioni note per esser presenti nel tessuto primario. La nostra analisi ha: *i)* mostrato una concordanza fra cfDNA e tessuto primario in circa il 67% dei pazienti e in particolare la tecnica NGS, basata sull'uso di sequenze specifiche (dette sequenze *barcode*), si è dimostrata migliore rispetto alla metodica NGS convenzionale "Amplicon-based NGS", consentendoci di identificare mutazioni con frequenza inferiore all'1%; i*i)* dimostrato che la tecnica dPCR piò essere utilizzata per monitorare tali mutazioni nel corso della terapia; *iii)* confermato il ruolo di biomarcatore di risposta molecolare precoce del cfDNA per i pazienti con HGSOC esplicato attraverso lo studio di mutazioni a carico di *TP53*, presenti al momento della progressione di malattia, ma non durante il trattamento chemioterapico.

Anche se il cfDNA rispecchia l'eterogeneità del tumore da cui ha origine, tale biomarcatore non fornisce informazioni sull'espressione genica e sulla componente cellulare proteica. Tale limite può oggi esser superato attraverso lo studio delle vescicole extracellulari (EVs). Si tratta di vescicole microscopiche prodotte da molte linee cellulari, compresi i tumori, la cui funzione

principale è legata ai meccanismi di comunicazione intercellulare espletata tramite il trasporto di informazioni molecolari dalla cellula di origine a quella bersaglio. A tal proposito, nel **Capitolo 5** abbiamo configurato un processo per l'isolamento e la caratterizzazione di RNA messaggero (mRNA) delle EVs (EV-RNA) in conformità alle linee guida MISEV2018 (Minimal information for studies of extracellular vesicles). Tale metodo è stato inizialmente testato su un modello cellulare per poi esser applicato sul plasma di 20 pazienti con carcinoma metastatico. L'analisi del profilo di espressione di 96 geni attraverso RT-qPCR e la percentuale di frequenza di mutazioni (VAF) a carico dei geni *PIK3CA*, *KRAS* e *BRAF* hanno rilevato una concordanza fra l'mRNA di origine cellulare e l'EV-RNA, suggerendo che le informazioni contenute in tali vescicole mimano quelle presenti nelle cellule da cui originano.

Alcune ricerche hanno dimostrato che per un isolamento efficiente ed affidabile di EVs integri e funzionali è indispensabile la scelta del corretto tubo di raccolta di sangue. Abbiamo pertanto voluto confrontare la quantità e la qualità di EVs ottenute da sangue raccolto in tre diversi tubi: EDTA, BCT e CellSave. Nonostante i tubi EDTA hanno mostrato risultati migliori, i nostri dati hanno dimostrato che il plasma deve essere raccolto entro e non oltre le 24 ore dal prelievo per ridurre al minimo il rischio di contaminazione leucocitaria, da noi studiata in termini di copie dei geni controllo of *GAPDH* and beta-2-microglobulina (*B2M*) di origine leucocitaria.

Il rilascio delle EVs rappresenta, per le cellule tumorali, un'opportunità che può favorire la progressione tumorale. Studi antecedenti al nostro hanno dimostrato che mutazioni specifiche del tumore posso essere identificate nell'RNA di origine vescicolare. Di conseguenza, abbiamo deciso di studiare le mutazioni note per esser presenti nel tessuto primario nell'EV-RNA attraverso la metodica dPCR. Nonostante la sensibilità del nostro metodo, tali mutazioni sono state identificate solo nel 15% dei casi. È ipotizzabile che questi risultati dipendano dalla scarsa quantità di plasma usato per isolare EVs (< 1 mL) e dalla mancanza di un metodo specifico per l'arricchimento della componente di EVs di origine tumorale.

E' noto che circa il 14-39% dei pazienti con carcinoma metastatico al seno (MBC) sviluppa resistenza alla terapia endocrina a causa della presenza di mutazioni nel gene che codifica per il recettore degli estrogeni (m*ESR1*). La maggior parte di queste mutazioni interessa una regione all'interno del dominio di legame del ligando di *ESR1* e consiste di sostituzioni nucleotidiche a carico degli amminoacidi in posizione 537 e 538, localizzati nell'elica 12 del recettore degli estrogeni. Di conseguenza, l'identificazione di tali mutazioni gioca un ruolo chiave nella scelta terapeutica di pazienti con MBC. Nonostante il cfDNA sia facilmente accessibile esso è spesso

presente a basse concentrazioni. A questo proposito, nel **Capitolo 6** abbiamo condotto uno studio sull'ottimizzazione del processo di isolamento del cfDNA/ctDNA e la caratterizzazione delle m*ESR1* da una piccola quantità (200 µL) di plasma di pazienti con MBC. Il cfDNA isolato con una quantità compresa tra 2-200 ng è stato inizialmente amplificato in modo target-specifico (uniplex o multiplex-targets) e poi analizzato mediante dPCR cui è stato applicato un valore soglia di VAF stringente dell'1%. I risultati ottenuti sono stati poi confrontati con quelli riscontrati mediante analisi NGS effettuata attraverso l'uso di un pannello contente 10 geni implicati nella tumorigenesi del cancro al seno, fra i quali *ESR1*. Il nostro studio ha confermato un arricchimento delle mutazioni di *ESR1* in pazienti MBC al momento della progressione di malattia piuttosto che prima dell'inizio della terapia (34.1% *vs* 3.9%).

Infine, diversi studi hanno mostrato che le fusioni geniche possono diventare una potenziale causa di sviluppo o resistenza alla terapia del tumore. Nel carcinoma mammario tali alterazioni sono principalmente rilevate nella forma primaria, in particolar modo se aggressiva, come per i fenotipi luminale B, il basal-like, e nei carcinomi mammari resistenti a terapia endocrina. Tali fusioni geniche avvengono tipicamente fra il gene *ESR1* e un gene partner, come *CCDC170, AKAP12* e *ARMT1.* Dal punto di vista funzionale, questi riarrangiamenti codificano per proteine costitutivamente attive in grado di promuovere la crescita tumorale. Pertanto, nel **Capitolo 7** abbiamo voluto analizzare questi riarrangiamenti e correlare la loro presenza alla prognosi di questo gruppo di pazienti. La nostra analisi ha rilevato che i riarrangiamenti *ESR1-CCDC170*: *i)* sono più frequenti in soggetti con carcinoma mammario *ESR1*+ piuttosto che in pazienti *ESR1*- (frequenza: 25.8% *vs* 2.3%); *ii)* interessano maggiormente l'esone 2 di *ESR1* ed gli esoni 2, 6 e 8 del locus genico *CCDC170* ; *iii)* possono essere espressi, seppur a bassi livelli, in tessuti "sani", per esempio fibroadenoma (tumore benigno), tumore in situ duttale (DCIS) e al livello del tessuto epiteliale mammario non tumorale di donne con cancro al seno; *iv)* correlano con una peggiore prognosi, in termini di sopravvivenza libera da malattia (DFS) e sopravvivenza globale (OS); *v)* non hanno un valore predittivo di risposta alla terapia endocrina.

In conclusione, in questa tesi abbiamo evidenziato la necessità di disporre di biomarcatori selettivi e specifici (cfDNA/ctDNA ed EVs) dotati di valore predittivo e prognostico per l'analisi di pazienti con malattia tumorale. Per il loro utilizzo, è tuttavia necessario standardizzare ed ottimizzare l'intero processo analitico, dalla raccolta del campione alla suaanalisi. Solo in un tale contesto, la biopsia liquida si potrà confermare come valido alleato per la caratterizzazione

molecolare di un tumore, la scelta dell'approccio terapeutico, nonché la predizione di ricaduta, sviluppo di metastasi e resistenza alla terapia.

Appendices

PhD Portfolio

Name PhD student	**Silvia Rita Vitale**
Erasmus MC Department	Medical Oncology
Research School	Postgraduate School Molecular Medicine (MolMed)
PhD Period	January 2016 to February 2021
Promoters	Prof. JWM Martens
Copromotor	Dr. M.P.H.M. Jansen

PhD Training

	Year	Workload (ECTS)
General academic and research skills		
Biomedical English Writing	2017	2.0
Follow-up Photoshop & Illustrator CC	2017	0.3
Scientific Integrity Course	2017	0.3
In-depth courses		
Basic and Translational Oncology	2017	1.8
Basic Course on 'R'	2017	1.8
Basic Introduction on SPSS course	2017	1.0
Naic digital-PCR System Training, Rotterdam	2017	0.3
Biomek i-series software for Beginner, Catania, Italy	2019	0.8
Poster and Oral Presentations		
11th EORTC Meeting, Rotterdam (*oral presentation*)	2016	0.5
San Antonio Breast Cancer Symposium, Texas, USA (*poster*)	2016	0.5
Annual Molecular Medicine Day 21th Medical Medicine Day (*poster*)	2017	0.5
Annual Molecular Medicine Day 21th Medical Medicine Day (*oral presentation)*	2017	0.5

	Year	Workload (ECTS)
Scientific Meeting, Medical Oncology, Erasmus MC, Rotterdam (*poster*)	2017	0.5
qPCR dPCR and NGS meeting, Munich, Germany (*poster*)	2017	0.5
JNI Scientific Lab meeting, Erasmus MC, Rotterdam (*oral presentation*)	2017	0.5
Netherlands Society for Extracellular Vesicles (NLSEV)(*poster*)	2017	0.5
San Antonio Breast Cancer Symposium, Texas, USA (*poster*)	2018	0.5
International conferences		
San Antonio Breast Cancer Symposium, Texas, USA	2016	1.5
San Antonio Breast Cancer Symposium, Texas, USA	2018	1.5
Seminars and meetings		
Real-time PCR Training, Utrecht	2016	0.3
Annual Molecular Medicine Day 20th Medical Medicine Day	2016	0.3
4th Daniel Den Hoed Day, Rotterdam	2016	0.3
11th EORTC Meeting, Rotterdam	2016	0.3
Interlab-meeting Rotterdam/Antwerp	2016	0.2
JNI Scientific Lab meeting, Erasmus MC, Rotterdam	2016 - 2017	2.0
Medical Oncology Meeting, Erasmus MC, Rotterdam	2016 - 2017	0.7
cfDNA meeting, Erasmus MC, Rotterdam	2016 - 2017	0.3
EV meeting, Erasmus MC, Rotterdam	2016 - 2017	0.3
Journal Club, Erasmus MC, Rotterdam	2016 - 2017	0.8
^{5th} Daniel Den Hoed Day, Rotterdam	2017	0.3
Annual Molecular Medicine Day 21th, Rotterdam	2017	0.3
Annual meeting cancer genomics, Utrecht	2017	0.3
qPCR dPCR and NGS meeting, Munich, Germany	2017	0.9
Netherlands Society for Extracellular Vesicles (NLSEV)	2017	0.3

	Year	Workload (ECTS)
NGS to NGO: Next Generation Sequencing to Next Generation Oncologists, Milan, Italy	2018	0.9
Journal club at Center of oncology and Experimental Hematology, Catania, Italy	2018 - 2020	0.8
Annual Molecular Medicine Day 23th, Rotterdam	2019	0.3
One day: confronto sulla terapia ormonale e le nuove terapie per il ca. mammario avanzato, Catania, Italy	2019	0.2
CML Academy Network, Catania, Italy	2019	0.3
Supervising practical's and bachelor's thesis		
Internship of student Marjolein Gerritsen	2017	3.0
Internship of student Fereshta Ataullah	2017	3.0

List of publications

1. Stagno F, Vigneri P, Cupri A, **Vitale SR,** Di Raimondo F. Diagnosis of blastic phase of chronic myeloid leukaemia. Acta Haematol. 2012;127(4):198

2. Stagno F, Vigneri P, Cupri A, **Vitale SR**, Di Raimondo F. Infliximab therapy in hematologic malignancies: handle with care. Haematologica. 2012 Aug;97(8):e26.

3. Manzella L, Tirrò E, Pennisi MS, Massimino M, Stella S, Romano C, **Vitale SR**, Vigneri P. Role of Interferon regulatory factors in Chronic Myeloid Leukemia. Curr Cancer Drug Targets. 2016 Jan 4

4. van Dessel LF, Beije N, Helmijr JC, **Vitale SR,** Kraan J, Look MP, de Wit R, Sleijfer S, Jansen MP, Martens JW, Lolkema MP. Application of circulating tumor DNA in prospective clinical oncology trials - standardization of preanalytical conditions. Mol Oncol. 2017 Mar;11(3):295-304. doi: 10.1002/1878-0261.12037. Epub 2017 Feb 22.

5. Beije N, Sieuwerts AM, Kraan J, Van NM, Wendy O, **Vitale SR**, van der Vlugt-Daane M, Dirix LY, Brouwer A, Hamberg P, de Jongh FE, Jager A, Seynaeve CM, Jansen MPHM, Foekens JA, Martens JWM, Sleijfer S.. Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients. Mol Oncol. 2017 Oct 24. doi: 10.1002/1878-0261.12147.

6. Massimino M, Tirrò E, Stella S, Frasca F, Vella V, Sciacca L, Pennisi MS, **Vitale SR**, Puma A, Romano C, Manzella L. Effect of combined epigenetic treatments and ectopic NIS expression on undifferentiated thyroid cancer cells. Anticancer Res. 2018 Dec;38(12):6653-6662. doi: 10.21873/anticanres.13032.

7. Pennisi MS*, Stella S*, **Vitale SR,*** Puma A, Di Gregorio S, Romano C, Tirrò E, Massimino M, Antolino A, Siragusa S, Mannina D, Impera S, Musolino C, Mineo G, Martino B, Zammit V, Di Raimondo F, Manzella L, Stagno F, Vigneri P. BCR-ABL1 Doubling-Times and Halving-Times May Predict CML Response to Tyrosine Kinase Inhibitors. Front Oncol. 2019 Aug 13;9:764. Doi: 10.3389/fonc.2019.00764.

8. **Vitale SR***, Sieuwerts AM*, Beije N, Kraan J, Angus L, Mostert B, Reijm EA, Van NM, van Marion R, Dirix LY, Hamberg P, de Jongh FE, Jager A, Foekens JA, Vigneri P, Sleijfer S, Jansen MPHM, Martens JWM. An Optimized Workflow to Evaluate Estrogen Receptor Gene Mutations in Small Amounts of Cell-Free DNA. J Mol Diagn. 2019 Jan;21(1):123-137. doi: 10.1016/j.jmoldx.2018.08.010. Epub 2018 Oct 6.

9. Tirrò E, Massimino M, Romano C, Pennisi MS, Stella S, **Vitale SR**, Fidilio A, Manzella L, Parrinello NL, Stagno F, Palumbo GA, La Cava P, Romano A, Di Raimondo F, Vigneri PG. Chk1 inhibition restores inotuzumab ozogamicin citotoxicity in CD22-positive cells expressing mutant p53. Front Oncol. 2019 Feb 18;9:57. doi: 10.3389/fonc.2019.00057.

10. Tirrò E, Massimino M, Stella S, Zammit V, Consoli ML, Pennisi MS, **Vitale SR**, Romano C, Pirosa MC, Martino E, DI Gregorio S, Puma A, DI Raimondo F, Manzella L, Stagno F. Efficacy of Nilotinib in a CML Patient Expressing the Three-way Complex Variant Translocation t(2;9;22). Anticancer Res. 2019 Jul;39(7):3893-3899. doi: 10.21873/anticanres.13540.

11. Stella S, Tirró E, Massimino M, **Vitale SR**, Russo S, Pennisi MS, Puma A, Romano C, DI Gregorio S, Innao V, Stagno F, DI Raimondo F, Musolino C, Manzella L. Successful management of a pregnant patient with chronic myeloid leukemia receiving standard dose imatinib. In Vivo. 2019 Sep-Oct;33(5):1593-1598. doi: 10.21873/invivo.11641.

12. Stella S, Massimino M, Tirrò E, **Vitale SR**, Scalise L, Leotta S, Pennisi MS, Puma A, Romano C, Stagno F, Sapienza G, Milone G, Manzella L. B-ALL relapses after autologous stem cell transplantation associated with a shift from e1a2 to e14a2 BCR-ABL transcripts: A case report. Anticancer Res. Jan;39(1):431-435. doi: 10.21873/anticanres.13130.

13. Tirrò E, Stella S, Massimino M, Zammit V, Pennisi MS, **Vitale SR**, Romano C, Di Gregorio S, Puma A, Di Raimondo F, Stagno F, Manzella L. Colony-Forming Cell Assay Detecting the Co-Expression of JAK2V617F and BCR-ABL1 in the Same Clone: A Case Report. Acta Haematol. 2019;141(4):261-267. doi: 10.1159/000496821. Epub 2019 Apr 9.

14. Massimino M, Stella S, Tirrò E, Consoli ML, Pennisi MS, Puma A, **Vitale SR**, Romano C, Zammit V, Stagno F, DI Raimondo F, Manzella L. Efficacy of dasatinib in a very elderly CML patient expressing a rare E13A3 BCR-ABL1 fusion transcript: A case report. Anticancer Res. 2019 Jul;39(7):3949-3954. doi: 10.21873/anticanres.13548.

15. Stella S., Massimino M., Tirrò E., **Vitale S.R**., Accurso V., Puma A., Pennisi MS., Di Gregorio S., Romano C, Di Raimondo F., Siragusa S., Manzella L. Detection and clinical implications of a novel BCR-ABL1 E12A2 insertion/deletion in a CML patient expressing the E13A2 isoform. Anticancer Reaserch. Volume 39, Issue 12, 2019, Pages 6965-6970. Doi: 10.21873/anticanres.13918. Anticancer Res. 2019 Jan;39(1):431-435. doi: 10.21873/anticanres.13130.

16. Lisanne F. van Dessel*, **Silvia R. Vitale***, Jean C. A. Helmijr, Saskia M. Wilting, Michelle van der Vlugt-Daane, Esther Oomen-de Hoop, Stefan Sleijfer, John W. M. Martens, Maurice P. H. M. Jansen, Martijn P. Lolkema. High-throughput isolation of circulating tumor DNA: a comparison of automated platforms. Mol Oncol. 2019 Feb; 13(2): 392–402. Published online 2018 Dec 22. doi: 10.1002/1878-0261.12415.

17.Stefania Stella, Valentina Zammit, **Silvia Rita Vitale**, Maria Stella Pennisi, Michele Massimino, Elena Tirrò, Stefano Forte, Antonio Spitaleri, Agostino Antolino, Sergio Siragusa, Vincenzo Accurso, Donato Mannina, Stefana Impera, Caterina Musolino, Sabina Russo, Alessandra Malato, Giuseppe Mineo, Maurizio Musso, Ferdinando Porretto, Bruno Martino, Francesco Di Raimondo, Livia Manzella, Paolo Vigneri, Fabio Stagno. Clinical Implications of Discordant Early Molecular

Responses in CML Patients Treated with Imatinib. Int J Mol Sci. 2019 May; 20(9): 2226. Published online 2019 May 6. doi: 10.3390/ijms20092226.

18. Livia Manzella, Michele Massimino, Stefania Stella, Elena Tirrò, Maria Stella Pennisi, Federica Martorana, Gianmarco Motta, **Silvia Rita Vitale**, Adriana Puma, Chiara Romano, Sandra Di Gregorio, Marco Russo, Pasqualino Malandrino, Paolo Vigneri. Activation of the IGF Axis in Thyroid Cancer: Implications for Tumorigenesis and Treatment. Int J Mol Sci. 2019 Jul; 20(13): 3258. Published online 2019 Jul 2. doi: 10.3390/ijms20133258.

19. Elena Tirrò, Federica Martorana, Chiara Romano, **Silvia Rita Vitale**, Gianmarco Motta, Sandra Di Gregorio, Michele Massimino, Maria Stella Pennisi, Stefania Stella, Adriana Puma, Fiorenza Gianì, Marco Russo, Livia Manzella, Paolo Vigneri. Molecular alterations in thyroid cancer: From bench to clinical practice. Genes. Volume 10, Issue 9, September 2019, Article number 709. Doi: 10.3390/genes10090709

20. Massimino M., Stella S., Tirrò E., Consoli M.L., Pennisi M.S., Puma A., **Vitale S.R.,** Romano C., Zammit V., Stagno F., Di Raimondo F., Manzella L. Rapid decline of philadelphia-positive metaphases after nilotinib treatment in a CML patient expressing a rare e14a3 BCR-ABL1 fusion transcript: A case report. Oncology Letters. Volume 18, Issue 3, September 2019, Pages 2648-2653. Doi: 10.3892/ol.2019.10558

21. Stella S., Gottardi E.M., Favout, V., Gonzalez, E.B, Errichiello S**., Vitale SR**., Fava C., Luciano L., Stagno F., Grimaldi F., Pironi L., Simarro C.S., Vigneri P., Izzo B. The Q-LAMP method represents a valid and rapid alternative for the detection of the BCR-ABL1 rearrangement in Philadelphia-positive Leukemias. International Journal of Molecular Sciences.Volume 20, Issue 24, 2 December 2019, Article number 6106. Doi: 10.3390/ijms20246106

22. Massimino M., Tirrò E., Stella S., Pennisi M.S., **Vitale S.R**., Puma A., Romano C., Di Gregorio S., Romeo M.A., Di Raimondo F., Manzella L. Targeting BCL-2 as a therapeutic strategy for primary p210BCR-ABL1-positive B-ALL cells. In Vivo. Volume 34, Issue 2, 2020, Pages 511-516. Doi: 10.21873/invivo.11802.

23. **Vitale S.R.,** Floris H Groenendijk, Ronald van Marion, Corine M Beaufort, Jean C Helmijr, Hendrikus Jan Dubbink, Winand N M Dinjens, Patricia C Ewing-Graham, Ramon Smolders, Helena C van Doorn, Ingrid A Boere, Els M J J Berns, Jozien Helleman, Maurice P H M Jansen. TP53 Mutations in Serum Circulating Cell-Free Tumor DNA As Longitudinal Biomarker for High-Grade Serous Ovarian Cancer. Biomolecules, Volume 10, Issue 3, March 2020, Article number 415. Doi: 10.3390/biom10030415.

24. Manzella L., Tirrò E., **Vitale S.R**, Puma A., Consoli ML., Tambè L., Pennisi MS., DI Gregorio S., Romano C., Tomarchio C, Di Raimondo F., Stagno F. Optimal Response in a Patient With CML Expressing BCR-ABL1 E6A2 Fusion Transcript With Nilotinib Therapy: A Case Report. In Vivo. 2020;34(3):1481-1486. doi:10.21873/invivo.11933.

25. Massimino M, Stella S, Tirrò E, Pennisi MS., **Vitale S.R**, Puma A., Romano C., Di Gregorio S., Tomarchio C., DI Raimondo F., Manzella L. ABL1-Directed Inhibitors for CML: Efficacy, Resistance and Future Perspectives. Anticancer Res. 2020;40(5):2457-2465. doi:10.21873/anticanres.14215.

26. Stella S, Massimino M, Manzella L, Pennisi MS, Tirrò E, Romano C, **Vitale SR**, Puma A, Tomarchio C, Gregorio SD, Palumbo GA, Vigneri P. Molecular Pathogenesis and Treatment Perspectives for Hypereosinophilia and Hypereosinophilic Syndromes. Int J Mol Sci. 2021 Jan 6;22(2):486. doi: 10.3390/ijms22020486. PMID: 33418988; PMCID: PMC7825323.

27. Tirrò E, Massimino M, Romano C, Martorana F, Pennisi MS, Stella S, Pavone G, Di Gregorio S, Puma A, Tomarchio C, **Vitale SR**, Manzella L, Vigneri P. Prognostic and Therapeutic Roles of the Insulin Growth Factor System in Glioblastoma. Front Oncol. 2021 Feb 2;10:612385. doi: 10.3389/fonc.2020.612385. PMID: 33604294; PMCID: PMC7885861.

28. Martorana F, Motta G, Pavone G, Motta L, Stella S, **Vitale SR,** Manzella L, Vigneri P. AKT Inhibitors: New Weapons in the Fight Against Breast Cancer? Front Pharmacol. 2021 Apr 29;12:662232. doi: 10.3389/fphar.2021.662232. PMID: 33995085; PMCID: PMC8118639.

29. **Vitale SR**, * Helmijr JA*, Gerritsen M, Coban H, van Dessel LF, Beije N, van der Vlugt-Daane M, Vigneri P, Sieuwerts AM, Dits N, van Royen ME, Jenster G, Sleijfer S, Lolkema M, Martens JWM, Jansen MPHM. Detection of tumor-derived extracellular vesicles in plasma from patients with solid cancer. BMC Cancer. 2021 Mar 24;21(1):315. doi: 10.1186/s12885-021-08007-z. PMID: 33761899; PMCID: PMC7992353.

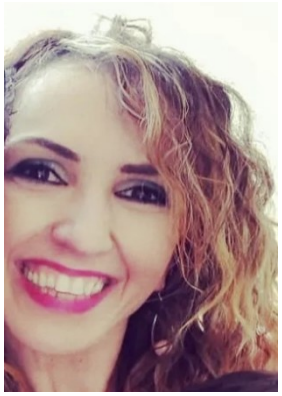
30. **Vitale SR**, Martorana F, Stella S, Motta G, Inzerilli N, Massimino M, Tirrò E, Manzella L, Vigneri P. PI3K inhibition in breast cancer: Identifying and overcoming different flavors of resistance. Crit Rev Oncol Hematol. 2021 Jun;162:103334. doi: 10.1016/j.critrevonc.2021.103334. Epub 2021 Apr 15. PMID: 33865994.

31.Romano C, Martorana F, Pennisi MS, Stella S, Massimino M, Tirrò E, **Vitale SR,** Di Gregorio S, Puma A, Tomarchio C, Manzella L. Opportunities and Challenges of Liquid Biopsy in Thyroid Cancer. Int J Mol Sci. 2021 Jul 19;22(14):7707. doi: 10.3390/ijms22147707. PMID: 34299334; PMCID: PMC8303548.

32.Massimino M, Tirrò E, Stella S, Manzella L, Pennisi MS, Romano C, **Vitale SR**, Puma A, Tomarchio C, Di Gregorio S, Antolino A, Di Raimondo F, Vigneri P. Impact of the Breakpoint Region on the Leukemogenic Potential and the TKI Responsiveness of Atypical *BCR-ABL1* Transcripts. Front Pharmacol. 2021 Jun 30;12:669469. doi: 10.3389/fphar.2021.669469. PMID: 34276365; PMCID: PMC8277938.

33. Stella S, **Vitale SR**,*Massimino M, Puma A, Tomarchio C, Pennisi MS, Tirrò E, Romano C, Martorana F, Stagno F, Di Raimondo F, Manzella L. A Novel System for Semiautomatic Sample Processing in Chronic Myeloid Leukaemia: Increasing Throughput without Impacting on Molecular Monitoring at Time of SARS-CoV-2 Pandemic. Diagnostics (Basel). 2021 Aug 20;11(8):1502. doi: 10.3390/diagnostics11081502.

Curriculum Vitae



Silvia Rita Vitale was born on April 14th 1985 in Catania (Sicily, Italy). In July 2009, she obtained her degree in Cellular and Molecular Biology from the University of Catania. During University Silvia was also a student at the "*Scuola Superiore di Catania*", the research-oriented honors program of University of Catania, where she passed additional exams next to the normal curriculum. In September 2009 she started working at the Oncology laboratory of the Department of Clinical and Experimental Medicine (University of Catania) supervised by Prof. dr. P. Vigneri and mentored by dr. Stefania Stella. In this laboratory, she has developed a solid background in molecular biology and has been involved in the expansion of the Molecular Diagnostic Unit, which is in the charge of the molecular monitoring of patients with Chronic Myeloid Leukaemia (CML) and Philadelphia negative myeloproliferative disorders, for both Sicily and Calabria. After graduation, she also attended the post-lauream course (5 years duration) for a specialization in Clinical Biochemistry and Molecular Pathology, and she has obtained specialization in July 2017.

Then, her research interests were directed towards next generation sequencing (NGS) and liquid biopsies for solid tumors. In this context, she moved to the laboratory of Medical Oncology at the Erasmus Medical Center (Rotterdam, Netherlands) at the beginning of January 2016, to gain knowledge and skills on liquid biopsy analysis. During the period spent in Rotterdam, she helped to optimize the workflow for the analyses of *ESR1* mutations in cfDNA of metastatic breast cancer patients. She also presented results of this work at a poster presentation in December 2016 at the San Antonio Breast Cancer Symposium (Texas, USA). Moreover, she was involved in the optimization of the workflow for the analysis of extracellular vesicles (EVs) released from cancer patients. Moreover, during her period in Rotterdam, she was actively involved in various projects. Her activities, as part of these projects, concerned not only continuing with techniques already established in the molecular oncology laboratory, but also setting up new techniques, which are now routinely used by other members of that group. In 2016-2017 Silvia started her MSc/PhD program in the molecular oncology group under the supervision of Prof.dr. JWM Martens, and mentored by dr. MPH Jansen and dr. AM Sieuwerts. In 2018, she moved back to Sicily continuing her research collaboration with Rotterdam continued with a project focused on *ESR1* gene fusions in primary breast cancer. In fact, she presented results of this work at a poster presentation in December 2018 at the San Antonio Breast Cancer Symposium (Texas, USA). Her PhD research has led to five published scientific articles, four already published, and all of them as a first/co-first author.

In Sicily she is currently involved in molecular diagnostic for hematological and solid tumors (CML and Heredo-Familial Tumors) and in performing tests to address the toxicity of certain anticancer drugs (on *DPYD* gene for capecitabine and *UGT1A1* gene for irinotecan), as well as in projects about translational research for breast cancer.

Acknowledgments

Ed eccomi giunta alla fine di questo lungo percorso. Non avrei mai potuto immaginare di poter riuscire ad arrivare al traguardo, tanto inatteso quanto sperato.

Ricordo ancora quando per la prima volta, nel lontano 2015, parlai a Stefania del mio desiderio di fare un esperienza all'estero; temevo che non mi avrebbe sostenuta, anche perché andare via avrebbe gravato sul lavoro dei mei colleghi; invece, lei mi incoraggiò nel chiederlo al prof Vigneri il quale, con mia grande sorpresa, diede la sua approvazione... da li tutto ebbe inizio.

...Credo avesse incontrato per la prima volta il Prof. Sleijfer in un congresso in Francia, mi parlò di quanto fosse stato affascinato dalla sua attività di ricerca traslazionale e mi propose di andare a Rotterdam, all'Erasmus Medical Center, per imparare una nuova metodica.... mi propose l'Olanda! Ero timorosa… io che da sempre avevo preferito restare legata alla mia terra e alle mie abitudini, io che da sempre preferisco mettermi in dubbio piuttosto che avere fiducia in me stessa, io che da sempre temo il "nuovo" per paura di non essere all'altezza o semplicemente perché mi da più sicurezza rispetto a quanto ignoto. Eppure, nonostante tutti i dubbi, avevo bisogno di tentare, di mettermi alla prova, di uscire dalla routine, di buttarmi a capofitto in una esperienza, di non restare ancorata alla situazione in cui mi trovavo, di imparare, di tentare, di sbagliare, ma, soprattutto, avevo bisogno di crescere e di trovare la mia identità...

Da lì ricordo ancora la prima *call* su *Skype*, mi chiesero chi fossi e cosa facessi, mamma mia che ansia avevo di parlare a degli sconosciuti in una lingua sconosciuta, per me che la timidezza l'ha sempre fatta da padrona! E invece andò bene, ricordo la prima visita, ottobre 2015… la testa mi frullava dal sentire parlare tutti, e soprattutto tutto il giorno, in inglese...

Eppure cosi ebbe inizio…Era il 1 gennaio 2016, un aereo portò me, accompagnata dal mio allora fidanzato, ora unico ed immenso compagno di vita, in quello che sarebbe stato il mio universo per due anni.

Che dire…aneddoti ne avrei tantissimi da ricordare, dalla macchinetta che si mangiò la carta bancomat lasciandomi senza soldi, all'incendio scoppiato nell'appartamento sotto casa, al perdermi nelle strade di Rotterdam... Eppure, oggi, ogni piccolo gesto, ogni piccolo evento lo ricordo come una grande avventura che mi ha accompagnata alla scoperta di questa immensa terra ... e di me stessa.

E quindi *In primis* dico grazie a te Olanda, a te Rotterdam, immensa distesa di prati fioriti dai variopinti tulipani, dal colore celeste del cielo nelle soleggiate giornate di maggio, dal profumo inconfondibile di freschezza, dai colori di tramonti che ti tolgono il respiro, dai mulini a vento che rievocano atmosfere fiabesche, dal vento forte che ti impedisce di camminare, dalla fredda neve che ti rende bambino, dalla gioia che si ha nel vivere una città a misura d'uomo, dove tutto è ben organizzato, puntuale, dove occorre avere solo due gambe per passeggiare ed ammirare quanto ti circonda…i tuoi parchi, pieni di gente che assapora anche un timido raggio di sole facendo pic-nic, o si rallegra con drink di fronte ai tuoi *river*, gente che sa vivere a 360 gradi, lavora a pieno ritmo la settimana, ma riesce anche a dare spazio alla propria vita privata essendo sempre leali verso il prossimo.

E proprio questo che mi hai insegnato, la schiettezza dell'essere sempre sinceri, limpidi e trasparenti, nonché l'esigenza di trovare un equilibrio fra la vita lavorativa e privata, la fiducia che ti si da nel metterti alla prova, nell'affrontare nuove sfide, perché in fondo tutto è possibile, tutto è fattibile, non esistono gerarchie, non esiste anzianità, esiste solo meritocrazia, e se meriti, se vali, se sei in gamba e capace ti viene riconosciuto e ti si da la possibilità di farti strada e di risaltare, indipendentemente dal fatto che tu possa essere l'ultima arrivata.

Dico grazie, grazie a te Olanda ed alle splendide persone che ho conosciuto, grazie a te ho imparato cosa significa il reale rispetto delle capacità altrui, grazie a te ho imparato ad esser più sicura in me stessa.

Now, the first three people I want to thank are my supervisor and promotor Prof.dr. John Martens, my co-promotor Dr. Maurice Jansen and my mentor, great friend, Dr. Anieta Sieuwerts.

To Prof.dr. John Martens, I want to thank you for all your support, advices, help, interesting questions and feedback. Particularly, to accept me as your PhD student although it was not supposed to be when I started my adventure in Rotterdam. Many thanks to believe and support me, I hope I will handle the occasion and that you will be proud of me. Thanks to your huge network, I also managed to connect with other people, and I had the chance to participate to international meeting presenting my research and data. I will never forgot your support.

To Maurice, you were not only my supervisor, but also a nice friend for me. I was really happy and grateful to work with you. You definitely helped me to become stronger, not only in the lab but also in my daily life. You were always open to listen to my ideas and, thanks to you, I improved my skills. You always believe me and you were the first person that thought that, maybe, I could be able to be a PhD student at the Erasmus MC. I am really grateful to you, to the time trying to talk in Italian, to the time spent in the office and to your help to improve my English. We always have nice and interesting talks, not only about research but also about economy, news, politics, kids, food and much more. You never left me alone, also now that I'm in Italy. I really think that I'd never make my PhD thesis possible without your support.

To Anieta, my dear, great, dearly, kind Anieta. I simply have no words to thank you for being by my side all the time.

You made all of this possible. You passed your immense passion for research to me. You were for me a teacher, a supervisor, a guide but, mostly, you were a great friend… My best friend, as once you told me. I will never forgot our talks in Dordrecht planning new experiments or analyzing what we obtained, your tender patience in explaining new things or teaching the Dutch language, our walks, the sushi (you loved with small eggs), our drinks in the nice restaurant in front of the Erasmus bridge, the "movie night for women" (do you remember?...how funny it was!), the time spent together organizing a presentation for a talk, the poster presentation, our travels in San Antonio, or in Munich, our interminable chats about research, life andour experiences. I'm so grateful to have known you and to have had you in my life. You have always been present and loving with me; once I asked you about the maternity and you said that you would not have been able to do this, but I'm sure that you have it, as you were a great "mother" for me. I hope to succeed in life to have some of your immense skills.

I also want tothank Prof.dr. Stefan Sleijfer, to give me the opportunity to enter in his inter-national famous department, particularly in his laboratory of Medical Oncology. Thanks to him, I was able to interact with his amazing crew and staff.

I also want to thank the reading committee: Prof. S. van Laere, Prof. E. Schuuring and Dr. M. Honing for their valuable time spent to read this thesis and their suggestions to improve its value. Thanks also to the opposing members for having the great chance to have them during my thesis defense.

Now, I would want to thank my friends and colleagues…

First of all, thanks to Jean, you are very unique and amazing! I will never forgot the time spent together, enjoyed our experiments (we are perfectly synchronized, I started ones and you were able to continue it, and the other way around). I learned so much from you, the technical approaches, the calculation you

made by *Excel* (I think I will never be able to manage them), the lab experiments, how to think to analyzed the results and improve them. I also loved spent time with you writing outside in the park or in the hall of the Erasmus,talking about life spent time and how to take things easier, enjoying what we have and not what we will never have. You always was for me a great friend. I also want to thank you for accepting being one of my paranymphs.

Thanks to Rosita, you was extremely kind, and always helping me in the lab, in the office, with documents, resolving several types of issues, for always you was there supporting all kind, until the very last moment…up to now that I'm in Italy.
Thanks to Antoniette, Joan and Saskia. You were always friendly and helpful.

Thanks to Jozien. I want to thank you for the collaboration, and for letting me participating withyou in the *TP53* project.

Thanks to the PhD students, Lisanne, Lindsay, Inge, Nick, Marjoline! I want to thank them for making our time there much nicer and enjoyable. Particularly, dear Lisanne, thanks for the time spent together with our experiments. I enjoyed so much working with you.

And my dear Marie, thanks for your friendship and for our nice talks. You was a very great friend.

Thanks to all the people in lab, Vanya, Michelle, Corinne, Kirsten, Mieke, Mai, Anita, Wendy, Jaco and Marcel. I also learned a lot from all of you. Thanks you for accepting me and making me feel at home.

I am also so grateful to Peggy for all her support, and for teaching me the NGS analysis.

Jan, thanks also to you, you was always so friendly, willing to help and was nice sharing with you the apartment for two years. You are a very nice and kind person, I will never forgot your love for flowers and tulips.

Ma la mia esperienza non sarebbe stata possibile senza il supporto del Prof. Vigneri. Caro Prof., voglio ringraziarla per aver creduto in me, per essersi fidato e per avermi dato la possibilità di vivere un'esperienza, una delle esperienze migliori della mia vita, che mi ha dato la possibilità di crescere professionalmente e personalmente.

Un ringraziamento speciale va a Stefania, da sempre collega, amica, mamma, sorella... non credo bastino gli appellativi per poter descrivere quello che sei per me. Una mentore, una guida, un supporto, una insegnante, una confidente. Senza te credo che mi sarei arresa diverse volte. Mi hai sempre dato buoni consigli, spronato e sostenuto durante la mia esperienza a RTD (sia moralmente che professionalmente), e continui a farlo adesso. Riuscivi a consolarmi quando mi sentivo sola dicendomi: "dai, manca poco, rientri e poi staremo insieme per sempre!". Due anni di lontananza e non sentirli grazie alle nostre interminabili chat e chiamate. Ti dico grazie perché probabilmente, senza il tuo sostegno, non sarei mai partita e non sarebbe stata la stessa avventura.

Voglio ringraziare la mia piccola "sorella" Adriana, per essermi stata accanto durante i due anni con le nostre interminabili telefonate dove, raccontandoci un po' di tutto, riuscivi a farmi sentire a casa.

Ringrazio poi tutti i miei colleghi, Livia, Cristina, Michele, Maria Stella, Chiara, Elena, Giusi per avermi sostenuta in questo progetto.

Un grazie immenso va alla mia famiglia, Papà, Rossana, Peppe, Ori, Nonna, Zie, cugine ed alle mie amiche di sempre Sabrina, Erika… Grazie Papi per avermi insegnato a non mollare ed a mettermi in gioco, la tua

intraprendenza mi sarà sempre da esempio. Nonnina, mia tenera e dolce nonnina, dico grazie a te che quando ti dissi che sarei andata via per due anni f acessi la forte e mi dicessi che dovevo farlo per il mio futuro. Questa esperienza mi ha purtroppo “rubato” due anni con te, ma tu non mi hai mai fatto pesare l’assenza, anzi mi hai sempre spronata a migliorare e fare di più.

Dico grazie anche a te mamma, mio grande angelo, che da sempre mi guidi e sostieni...

E infine, non posso che ringraziare l’uomo della mia vita, che mi accetta giorno dopo giorno con tutti i miei interminabili difetti e paranoie. Che mi ha accompagnato in Olanda prendendomi per mano, per poi insegnarmi a camminare da sola, che mi sostiene, mi sprona, mi stimola a migliorarmi giorno dopo giorno, che mi ha insegnato a valutare tutto in modo critico, a non farmi scoraggiare, ma a prendere di petto la realtà, scorporandola, analizzandola fino a trovare la soluzione. Dico grazie a te amore mio, Marco, senza di te e senza il tuo sostegno in ogni mia decisione non sarei mai riuscita a conseguire questo traguardo. E, ovviamente dico grazie al nostro piccolo terremoto Doddo che con pazienza ha accettato di “condividere” la sua mamma con gli interminabili articoli per la stesura di questa tesi. Spero che il tempo impiegato riesca prima o poi a ripagare l’amore immenso che mi date.

Ed un grazie particolare va a me stessa, per averci creduto e, nonostante lo sconforto di alcuni momenti, aver reso questo traguardo possibile.